

KARI ANTILA

Volumetric Image Segmentation for Planning of Therapies

Application to dental implants and muscle tumors

KARI ANTILA

Volumetric Image Segmentation
for Planning of Therapies
Application to dental implants and muscle tumors

ACADEMIC DISSERTATION

To be presented, with the permission of
the Faculty of Information and Communication Sciences
of Tampere University,
for public discussion at Tampere University
on 29th January, at 12 o'clock.

ACADEMIC DISSERTATION

Tampere University, Faculty of Information and Communication Sciences
Finland

*Responsible
supervisor
and Custos*

Professor Emerita
Ulla Ruotsalainen
Tampere University
Finland

Pre-examiners

Professor Koen Van Leemput
Athinoula A. Martinos Center for
Biomedical Imaging
Harvard Medical School
United States

Professor Lasse Lennu
LUT University
Finland

Opponent

Professor Miika Nieminen
University of Oulu
Finland

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Copyright ©2021 author

Cover design: Roihu Inc.

ISBN 978-952-03-1822-2 (print)

ISBN 978-952-03-1823-9 (pdf)

ISSN 2489-9860 (print)

ISSN 2490-0028 (pdf)

<http://urn.fi/URN:ISBN:978-952-03-1823-9>

PunaMusta Oy – Yliopistopaino
Vantaa 2020

ESIPUHE

Pitkän puhteen jälkeen mielessä päällimmäisenä on kiitollisuus. Välitön kiitos kuuluu niin monille teille, jotka ovat olleet mukana haalimassa rahoitusta, toteuttamassa menetelmiä ja jakamassa julkaisemisen tuskaa. Tätä työtä tehdessä ensiarvoisen tärkeäksi osoittautui myös monenlainen epäsuora ja työn ulkopuolinen apu. Olen kovasti kiitollinen esimiehilleni, jotka onnistuivat järjestämään suojaa organisaatioturbulensseilta, hallinnollisesta kuormasta, vapaata tarpeen tullen ja kollegoille jotka ovat osallistuneet projektireen vetämiseen kun aisat ovat itseltä meinanneet tipahtaa. Toivon, että osaan tästä työstä jakaa ansiota myös kaikille teille, jotka olleet raahaamassa ruudun äärestä lenkille, rokkiakatemialle, kokkikouluun, ylös alas lumisia nyppylöitä, raamattua tulkitsemaan ja usein aktiviteettien päätteeksi joko Gore-Texit, Lycrat tai hupparit yllä tuopin ääreen. Arvostan kovasti teitä kaikkia, jotka ovat olleet mukana järjestämässä reissuja, tapahtumia, sietäneet ajoittaista aloitteellisuuden puutetta ja sään mukaan eläneitä keväisiä ja syksyisiä aikataulua.

Haluan antaa ison kiitoksen ohjaajalleni Prof. Ulla Ruotsalaiselle. Hänen rohkaisunsa, asiantuntemuksensa ja sitkeytensä oli minulle korvaamatonta. Olen työtä tehdessäni saanut nauttia erityisen hyvästä mentoroinnista ja tästä erityiskiitokset kuuluvat Dosentti Jyrki Lötjöselle, Prof. Mark van Gilsille ja Prof. Ilkka Korhoselle. Projekti ei olisi edennyt maaliin ilman Prof. Koen van Leemputin ja Prof. Lasse Lensun esitarkastukseen vaikeana aikana antamaa laadukasta panostaan. Kiitos myös vastaväittäjälleni Prof. Miika Nieminen jo haasteen vastaanottamisesta ja aiheeseen paneutumisesta. Artikkeleiden kirjoittajajoukon ulkopuolelta haluan mainita Anna-Leena Vuorisen ja Tiina Takalokastarin joiden vertaistuki oli työn kannalta kriittisinä hetkinä erittäin tervetullutta.

Väitöstyötä ovat tukeneet Business Finland (ent. Tekes) ja Teknologian Tutkimuskeskus VTT Oy ja Solita Oy projekteillaan Tilakuva, ATLAS, EDIFI, SalweImage ja IVVES. VTT ja Solita lahjoittivat myös anteliaasti työaikaa kirjan viimeistelyyn. Työtä ovat arvokkain apurahoin tukeneet Suomen Kulttuurirahaston Artturi ja Aina Heleniuksen rahasto, Instrumentariumin Tiedesäätiö ja Helsingin yliopiston Etelä-Pohjalainen Osakunta.

Lopuksi haluan antaa kiitokseni perheelleni, isä-Matille ja sisko-Hannelle tuesta myös haasteellisinkin hetkinä. Erityisesti haluan maita Tuija Elon ja kiittää hänen työstään perheen ja itseni apuna. Setäni Esa on ollut minulle avuksi monin tavoin myös, kuten pitkiä tunteja toista Valtraa komentamalla.

Kuluva vuosi jäänee mieleen vähän muustakin kuin tämän työn niputtamisesta. Etätöiden ja eristäytymisen haasteellisena aikana iloa voi löytää ainakin musiikista ja sarkasmista. Upea esimerkki näiden kahden lajin ansiokkaasta yhdistämisestä tämän vuoden tunnelmiini löytyy Jarkko Martikaisen sanoituksesta kappaleeseen Raato raahautuu (Jarkko Martikainen & Luotetut miehet, Ruosterastaat-albumi, Sakara Rekords 2016).

Voiko repiä runoutta, siitä toivottomuudesta
joka sielua piirittää, heti kun olen valveilla

Voiko siivittää säveltä, täyttämään sitä tyhjyyttä
Voiko laululla parantaa, tiedä ei jollei yritä

No mä kokeilen, mä laulan sen
ja kun nyt laulan sen, niin huokaisen

Voi paska, eikö tää lopu koskaan
Multaa satelee hiljalleen, raato raahautuu huomiseen
ja huokaisee taas, voi paska

Kuka haluaa kuunnella, harmaantunutta laulua
Josta sattuma sävyt söi, potki nurkkaan ja mitätöi

Tiedä en mutta tiedän sen, nytkin moni muu ihminen
Mieltii kaikkensa loppua, vaikka kaikki on alussa

Kappaleen erityinen voima on sen kepeästi duurissa soivassa sävelessä.

Tampereella 7.12.2020

Kari Antila

ABSTRACT

Image segmentation, partitioning an image to consistent, meaningful segments, is a requirement for systematic analysis of its contents. Segmentation is used in medical diagnostics and as presented in this work, in treatment planning and therapy assessment. This work presents three robust and fast methods for two applications. The first two methods were designed facial bones to speed up dental implant planning workflows and the third for muscle tumors (uterine fibroids) to automate the mid- and post-treatment analysis of the results of ultrasound therapy.

Both facial bone structures and muscle tumors can take individual, even unpredictable shapes. The used volumetric (three-dimensional) imaging modalities may suffer from distortions and other types of losses of quality because of the constraints set by feasible exposure or available scanning time. A valid, clinical-grade segmentation method should solve the problem fast to minimize wait times in the therapy planning workflow or almost real time when used to update the plan during the therapy.

To meet these needs we first developed a method that is capable segmenting mandibles from Narrow-Beam Volumetric Tomography images. It works by deforming a pre-constructed surface model around the mandibular bone. Our requirements were later upgraded to include all visible facial bones in Cone-Beam Computed Tomography images. For this revised goal we developed a novel data-driven method that reconstructs facial bone surfaces from continuous patches and bridges over holes due to missing teeth or image distortions. When our target shifted from the mandibular bone to the muscle tumor segmentation from Magnetic Resonance images, we were able to carry over the core properties of the algorithm to the new problem successfully.

We verified the robustness of both facial bone and tumor segmentation with independent training and validation sets and found their accuracy to match other published work. The requirement for a very tight computing budget was reached with as fast as under a minute processing time per image volume.

TIIVISTELMÄ

Kuvien segmentointi, mikä tarkoittaa kuvien sisällön jaottelua yhtenäisiin ja merkityksellisiin alueisiin, on välttämätön edellytys kuvien sisällön järjestelmälliselle analyysille. Lääketieteessä segmentointia käytetään diagnosoinnissa ja kuten tässä työssä, hoitojen suunnittelussa ja niiden tulosten arvioinnissa. Tämä väitöstyö esittelee kolme tarkkaa, luotettavaa ja nopeaa segmentointimenetelmää, joista kaksi ensimmäistä kehitettiin hammasimplanttien suunnittelua ja kolmas kohdun lihaskasvainten ultraääniterapian aikaisten ja jälkeisten tulosten analyysiä varten.

Sekä kasvojen luurakenteet, että lihaskasvaimet voivat saada yksilöllisiä ja vaikeasti ennustettavia muotoja. Myös käytetyt kolmiulotteiset kuvantamismenetelmät kärsivät erilaisista häiriöistä ja muista laatuheikkenemistä, jotka voivat johtua esimerkiksi kuvaamiseen käytettävän ajan ja mahdollisen säteilyaltistuksen asettamista rajoitteista. Kliiniseen käyttöön soveltuvan segmentointimenetelmän tulisi toimia luotettavasti ja suorittaa tehtävä nopeasti jotta odotusajat hoidon suunnittelussa minimoituisivat tai lähes reaaliaikaisesti hoitosuunnitelmaa hoidon aikana päivitettäessä.

Näitä tarpeita varten kehitettiin ensin alaleuan segmentointimenetelmä kapean kulman tomografiakuvia varten. Menetelmä käyttää ennalta koostettua pintamallia, mikä sovitetaan leukaluun ympärille. Työn aikana menetelmän käyttötarkoitusta laajennettiin kasvojen muidenkin luiden segmentointiin kartiokeila-tomografiakuvista. Tämä käyttötarkoitus edellytti uudenlaisen, datalähtöisen menetelmän kehittämistä kasvojen luuston yhtenäisten osapintojen tunnistamiseen ja niiden yhdistämiseen kuvien häiriöitä ja kohteiden epäjatkuvuuksia sietäen. Tämän algoritmin pääperiaatteet olivat sovellettavissa myös myöhemmin tehtäväksi asetettuun magneettikuvattujen lihaskasvainten segmentointiongelmaan.

Kehitettyjen menetelmien luotettavuus testattiin erillisillä opetus- ja validointijoukoilla ja niiden tarkkuus todettiin vastaavan samankaltaisia julkaistuja menetelmiä. Menetelmät suunniteltiin täyttämään tiukat suoritusaikavaatimukset ja suoritusajoissa onnistuttiin pääsemään jopa alle minuuttiin per kolmiulotteinen kuva.

CONTENTS

1	Introduction	1
2	Background.....	5
2.1	Image segmentation, definition.....	5
2.2	Imaging modalities	7
2.2.1	Orthopantomographic tomography	7
2.2.2	Narrow-Beam Volumetric Tomography.....	9
2.2.3	Cone-Beam Computed Tomography	11
2.2.4	MR Imaging with High-Intensity Focused Ultrasound.....	12
2.3	Error metrics for a segmentation algorithm.....	13
2.4	Prior work	14
2.4.1	Facial bone segmentation	15
2.4.2	Image view synthetization.....	19
2.4.3	Muscle tumor segmentation	20
3	Materials and Methods	22
3.1	Materials	22
3.2	Image processing pipeline.....	22
3.3	Model building	24
3.4	Image pre-processing.....	28
3.4.1	Image filtering	28
3.4.2	Resampling to a uniform grid.....	29
3.4.3	Equalization	29
3.4.4	Thresholding.....	30
3.5	Segmentation	31
3.5.1	NBVT pipeline.....	33
3.5.2	CBCT pipeline.....	35
3.5.3	MR-HIFU pipeline.....	39
4	Results	41
5	Discussion.....	46
6	Conclusion.....	50
7	References.....	51

ABBREVIATIONS

AAM	Active appearance model
ACM	Active contour model
ART	Algebraic reconstruction technique
ASM	Active shape model
CT	Computerized tomography (or Computed tomography)
CBCT	Cone-beam computerized tomography
CNN	Convolutional neural networks
DC	Direct clustering
DOF	Degrees of freedom
DSC	Dice similarity coefficient
EM	Expectation maximization
EHR	Electronic health records
FBP	Filtered backprojection
FFD	Free-form deformations
FCM	Fuzzy C-means
FC	Fuzzy connectedness
FOV	Field of view
CLAHE	Contrast-limited adaptive histogram equalization
GLM	Generalized linear model
GLCM	Gray-level co-occurrence matrix
GLRLM	Gray-level run length matrix
GMM	Gaussian mixture-model
HIFU	High-intensity focused ultrasound
HU	Hounsfield unit
LAT	Limited angle tomography
MA	Multi-atlas
MAP	Maximum a posteriori
MAR	Metal artefact reduction
ML	Machine learning
MR-HIFU	Magnetic-resonance high-intensity focused ultrasound

MRI	Magnetic resonance imaging
MSD	Mean surface distance
NBVT	Narrow-beam volumetric tomography
NPV	Negative predictive value, Non-perfused volume
OAR	Organ at risk
OPT	Orthopantomographic tomography
PCA	Principal component analysis
PPV	Positive predictive value
PV	Perfused volume
RF	Random forests
ROI	Region of interest
SSM	Statistical shape model
VT	Volumetric tomography
VOI	Volume of interest
WDM	Wavelet density model
WT	Wavelet transform

ORIGINAL PUBLICATIONS

- Publication I Kari Antila, Mikko Lilja, Martti Kalke and Jyrki Lötjönen. 2006. Model-based segmentation of reconstructed dental x-ray volumes. *Proceedings of the 2006 IEEE International Conference on Image Processing (ICIP)*, p. 1933-1936.
- Publication II Mikko Lilja, Ville Vuorio, Kari Antila, Henri Setälä, Jorma Järnstedt and Mika Pollari. 2007. Automatic segmentation of the mandible from limited-angle dental x-ray tomography reconstructions. *Proceedings of the 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, p. 967-967.
- Publication III Kari Antila, Mikko Lilja, Martti Kalke and Jyrki Lötjönen. 2008. Automatic extraction of mandibular bone geometry for anatomy-based synthetization of radiographs. *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2008*, p. 490-493.
- Publication IV Kari Antila, Mikko Lilja and Martti Kalke. 2016. Segmentation of facial bone surfaces by patch growing from cone beam CT volumes. *Dentomaxillofacial Radiology*, 45(8), p. 20150435
- Publication V Kari Antila, Heikki J. Nieminen, Roberto Blanco Sequeiros and Gösta Ehnholm. 2014. Automatic segmentation for detecting uterine fibroid regions treated with MR-guided high intensity focused ultrasound (MR-HIFU). *Medical Physics*, 41(7), p. 073502

AUTHOR CONTRIBUTIONS

- Publication I Kari Antila developed the ROI model, segmentation pipeline and did all implementation and analysis work. He was the main author of the article.
- Publication II Kari Antila was the third author and took part in designing the Mandible model and the segmentation pipeline that was updated from Publication I. He took part in the writing and proofing work of the article. This article was included for being an integral part of the evolution of the segmentation methods for the dental imaging.
- Publication III Kari Antila was responsible of developing the segmentation method presented in Publication II further and did all implementation and analysis work. He was the main author of the article.
- Publication IV Kari Antila was the primary designer of the segmentation algorithm, the pipeline and the main author of the article.
- Publication V Kari Antila was the designer of the segmentation pipeline and the main author of the article.

1 INTRODUCTION

Since the discovery of X-rays in 1895, Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) in the early 1970s, the art and science of medical imaging has seen exponential growth in number of devices installed, images collected and thus data generated (Aiello et al., 2019). Managing and refining this mass of data to practical knowledge has emerged as a major challenge on its own (Dinov, 2016). The tools of machine learning (ML) have long been projected to take over the task of medical image analysis. Still, the workload of a radiologist has been estimated to have increased of up to 50% during the past five years (Alexander et al., 2019).

The most common use for medical images is in the making or supporting a diagnosis. Images are also needed when planning for a surgery and assessing the results of the operation. Utilization of images in these tasks have several bottlenecks. Unlike other biomarkers such as those in a standard blood test, raw images may not come with easily quantifiable and comparable measures. These measures need to be extracted, a task that typically requires some sort browsing, calculating or even drawing. Modern high-resolution, volumetric (three-dimensional) images may consist of hundreds of stacked image slices and several images may be taken during a session or a progression of a condition. If done manually, which is still often the case, this can amount to several minutes, even hours of tedious, repetitive work per patient.

Image-based planning is especially important in dental implantology and maxillofacial surgeries (Swennen et al., 2009). The planner needs to establish the right views to take the right measures, such as the length and width of the fitted tooth implant from the taken volumetric scans. Automatic generation of these views would speed up the planning workflow considerably. This could be achieved with an algorithm that fits a model including the orientations of the desired views to individual patient anatomy. The algorithm could be realized by segmenting the facial bones (mandible and maxilla) from the surrounding tissues and fitting the views along the segmented bone lines. The results would help in ensuring that the implants have the correct fit and avoiding costly damage to the surrounding tissues such as mandibular nerves and mucous membranes covering the maxillary sinuses (Çiftçi et

al., 2016). The primary anatomical areas of for interest in the implant planning are presented in Figure 1.

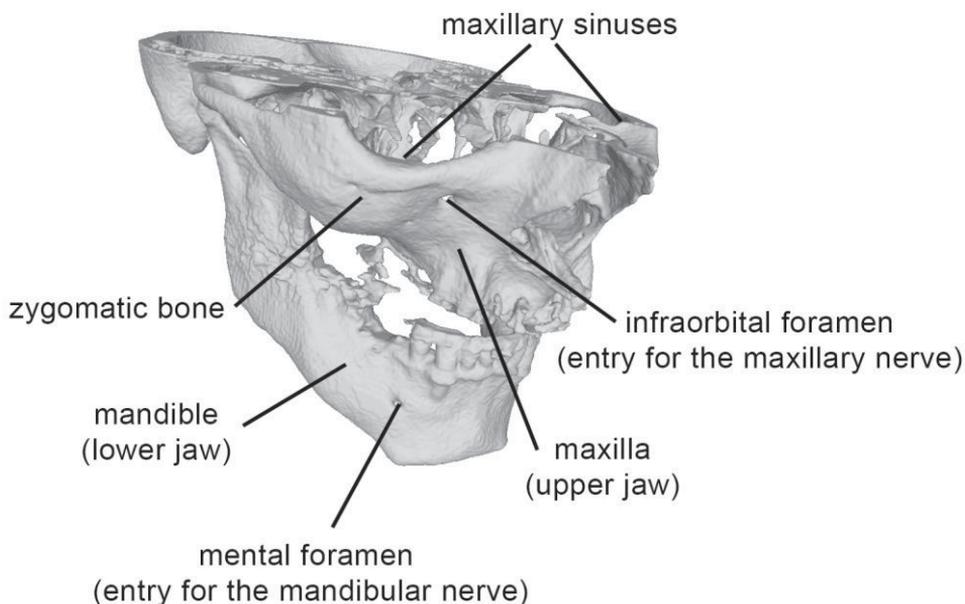


Figure 1. Facial bones and structures often relevant for dental implant and maxillofacial surgery planning.

The planning work in dental implantology and maxillofacial surgeries has traditionally been performed on Computerized tomography (CT) images (Bloch & Udupa, 1983). Although CT is able to give a very good contrast for the bone, there are two major disadvantages: CT usually requires a relatively high dose of ionizing x-ray radiation and suffers from a dramatic drop in image quality where there is metal in the view. Increased awareness of the risks of using ionizing radiation have been mitigated by developing new imaging modalities such as Narrow-beam volumetric tomography (NBVT) (Cederlund et al., 2014) and Cone-beam computerized tomography (CBCT) (M. Loubele et al., 2009). Despite the evolution of these methods, the severe distortions caused by metal (fillings, braces, implants) are still very common (Chindasombatjaroen et al., 2011), (Nardi et al., 2015). The reduction of dose in CBCT also introduces more noise, slightly poorer contrast and a range of other characteristic image distortions (M. Loubele et al., 2008), (Liang et al., 2010) when compared to CT.

The primary objective of this work was to develop a robust, fast and accurate segmentation method for facial bones in NBVT and CBCT. This method was required to tolerate noise and artefacts that are characteristic for low-dose, volumetric dental images. At the beginning of this study only a handful of dental segmentation methods had been published and none of those would perform reliably with the image material we had at hand. Our first segmentation method was based on fitting pre-made surface model around the mandible only. The second method was developed for the whole exterior of the face and was based on searching consistent surface patches on the facial bones and connecting them to larger surfaces. An example of a segmentation result and the orthogonal view generated is presented in Figure 2.

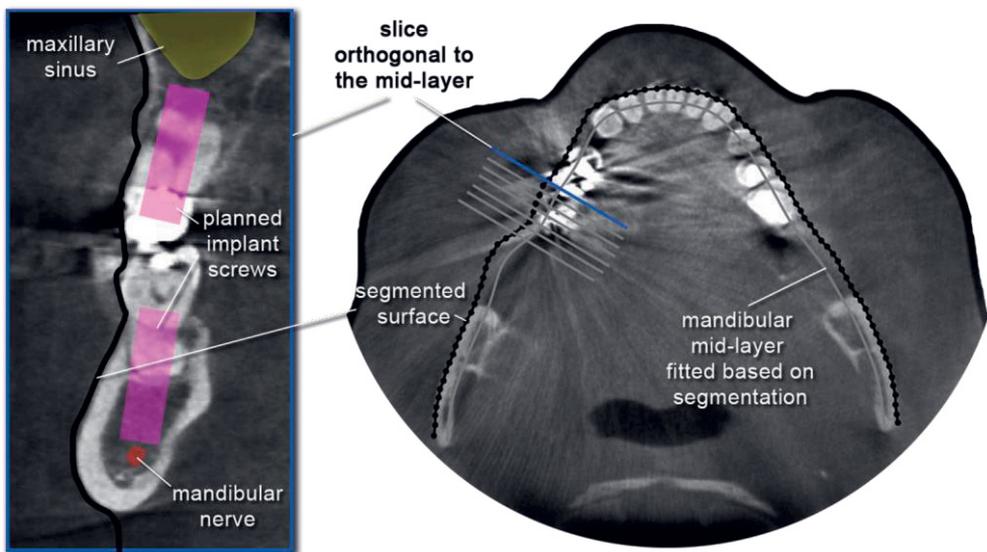


Figure 2. Segmentation result for CBCT. An orthogonal view (marked with blue segment on the right) is generated on the presumed implant location based on the segmentation. The implant's (in magenta on the left image) depth needs to be correct for it to attach firmly and not to damage the mandibular nerve (in red on the left) or the maxillary sinus (in yellow on the left).

Imaging can be a valuable and sometimes the only direct, non-invasive tool for the assessment of the success of an operation. Magnetic resonance-guided high-intensity focused ultrasound (MR-HIFU) (Kennedy, 2005), (Elhelf et al., 2018) is a form of therapy where the target, for example a muscle tumor, is given just enough heat to destroy the tumorous tissue without harming the surrounding organs. After, or even during the therapy, it is necessary to measure how much of the tumor volume had

been heated to the state of necrosis (cell death). The untreated tissue that has active blood flow can be enhanced with a Gadolinium showing bright when the treated, necrosed tissue appears dark in MR images taken repeatedly during the therapy. The volumes of the Gadolinium-perfused (untreated) and the non-perfused (treated, necrosed) regions are needed for a reliable estimate of the long-term success of the therapy (Lénárd et al., 2008). Segmentation of these volumes is generally challenging since both areas can take complex and arbitrary shapes.

The secondary objective of this work was set to produce an accurate segmentation method capable estimating these volumes (Figure 3) in post-operative images, so that the decisions whether to continue the treatment could be made based on the best and latest evidence.

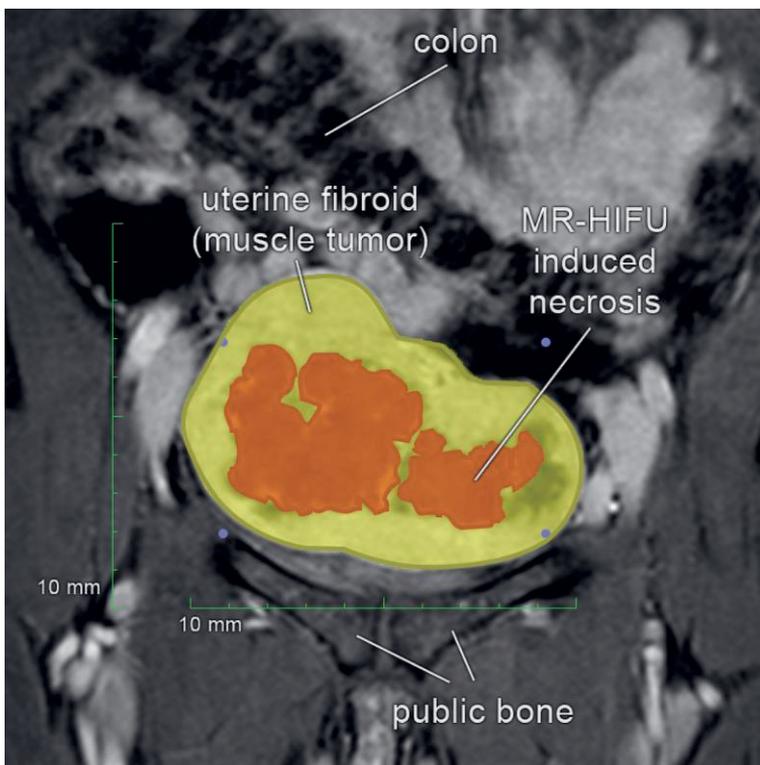


Figure 3. Segmentation result for a uterine fibroid after HIFU-therapy on an MR image. The tumor is shown in yellow and the necrosis inside the tumor in brown.

2 BACKGROUND

2.1 Image segmentation, definition

The challenge of automatic detection, recognition and interpretation of objects of interest in images and video has seen considerable research effort and investment over the history of imaging. The potential value of practical solutions to the problem has been recognized and realized in numerous areas of science and engineering including machine vision (Beyerer et al., 2015), surveying of natural resources (Unsalan & Boyer, 2013), security (AL-Allaf, 2014) and medical imaging. The term image segmentation usually refers to partitioning the contents of an image to segments that share some consistent, common property and is of special interest in the given application. The wide usage of image analysis techniques has led to inconsistent terminology. The following presents a consensus of the segmentation-related terms with the help of an example.

Automated vehicles need capabilities of analyzing a surrounding scenery (Figure 4). A bare minimum would be to have functionality to point out potential objects that could help in navigating and avoiding obstacles. Thus, a robust *object detection* algorithm would enable functions like automatic braking or a parking assistant. More advanced vehicle automation would benefit from *object recognition*, where the objects around the vehicle would be assigned pre-defined classes such as signs, cars or persons by the characteristic shape, color or texture they show. This would give the application controlling the vehicle a better capability to analyze the dynamic nature of the surroundings (which objects are likely to stay still, which will move and at what speed) and thus a preemptive vehicle control. A step deeper analysis of the surroundings could benefit from partitioning the scene into separate objects along their edges (borders, contours), for example to estimate where the lanes and sidewalks lay. The task of partitioning the whole scene this way is usually called *semantic segmentation*. A specific object class could be further separated from other classes and the background with *instance segmentation*. When semantic segmentation usually partitions the whole scene and treats individual objects belonging to a class as a single entity, instance segmentation may be trained to handle distinct instances of the class as separate entities (Figure 4 (c) vs. (d)).



Figure 4. Object detection (a), object recognition (b), semantic segmentation (c) and instance segmentation (d).

In the context of this work, and biomedical imaging in general, the term *segmentation* usually refers to *instance segmentation* since the end goal is typically a rather detailed, quantitative analysis of objects of a specific class. In a general proper, statistical analysis of medical images is possible only after the regions of interest (ROIs)

(usually anatomical structures such as a tumor) have been correctly segmented along their contours and identified.

2.2 Imaging modalities

2.2.1 Orthopantomographic tomography

Orthopantomographic tomography (OPT) or panoramic radiography is an imaging modality developed for creating wide, panoramic views of the dentistry in a single x-ray scan. The core techniques and methodology of OPT were developed from the 1940s to 1960s (Paatero, 1961). Since its introduction, OPT has remained popular due to its speed, resolution and capability of capturing all teeth and other adjacent structures such as the mandibular nerve canals in a single, wide scan.

An OPT scan is taken by rotating a C-shaped arm is moved around patient's head in a carefully designed pattern so that the layer around the jaws will appear sharp and the objects in the background such as the spine will be blurred. (Figure 5).

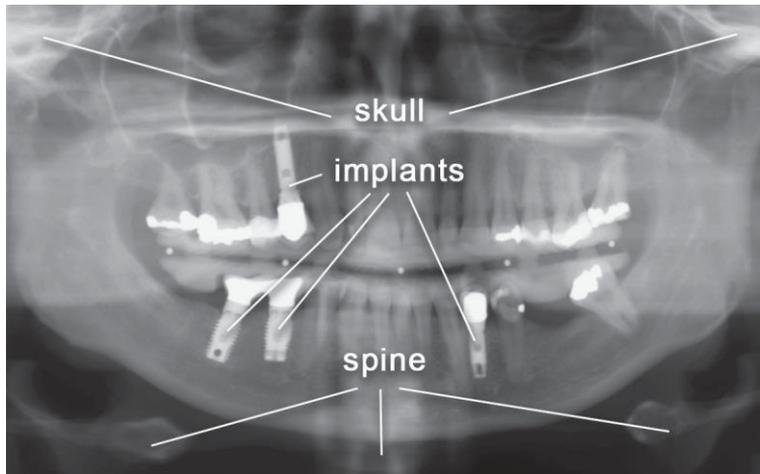


Figure 5. A typical OPT (panoramic) radiograph that shows all teeth in a single, sharp scan. However, it is not possible to conclusively determine the relative distances of objects in the depth orientation. The scan also suffers from shadow-like superposition of objects such as the skull and the spine. Modified from source material for the Publication III.

This form of tomography is sometime called “conventional tomography” (Littleton & Durizch Littleton, 1996) or “focal plane tomography”.

Since the curved tomographic plane of focus in OPT is fixed to the scanner, the correct placement of the patient is critical for acquiring the focus only on the jaws and not on the background. Even though the background items may appear less sharp, they will still be visible. Misplaced objects may appear larger (magnified) and cause errors in measurements. Finally, since OPT does not have depth information, the relative distance of objects such as teeth roots and nerves cannot be conclusively determined by OPT only.

This work does not involve segmenting or analyzing the OPT scans directly. However, the main motivation for Publications I-IV was to generate views for the planning of implants. Implant planning remains a popular use case for OPT. With the results from Publications I-IV it is possible to use a single volumetric scan like CBCT for synthesizing OPT-like views with patient-specific geometry and freely adjustable imaging parameters such as size, orientation, position and depth (Figure 6 and Figure 7). This addresses the problems of superimposition, artificial magnification and lack of depth information in traditional OPTs. Synthetization of OPTs from CBCT scans is beneficial also because one does not need to scan the subject multiple times. This limits the exposure, time and equipment required.



Figure 6. An OPT-like panoramic view synthesized from a CBCT scan. Since the geometry and depth of the reconstruction can be freely adjusted or fitted to individual patient anatomy, the scan does not suffer from superposition from the skull, spine or other structures. From Publication III. © IEEE 2008.



Figure 7. Visualization of a synthesized panoramic view (similar to Figure 6) over rendered bone surfaces.

2.2.2 Narrow-Beam Volumetric Tomography

Narrow Beam Volumetric Tomography is a modality where x-ray volumes can be reconstructed with regular OPT scanner hardware (Cederlund et al., 2014). NBVT was designed primarily to extend the capabilities of OPT scanners so that it is possible to capture also limited depth information for example in tracing roots of teeth in relation to the mandibular nerve to avoid damage.

NBVT is implemented by configuring a scanner to take a series of projection scans around the target with its line-type detector. However, the mechanics developed for an ordinary panoramic scanner do not allow capturing of projections from full 180 degrees around the target (Figure 8). Reconstructing an image with limited (incomplete) information do not yield perfect three-dimensional volumes and edges with tangential orientation outside the angle of view (AOV) cannot be reconstructed from the projections precisely (Quinto, 2017).

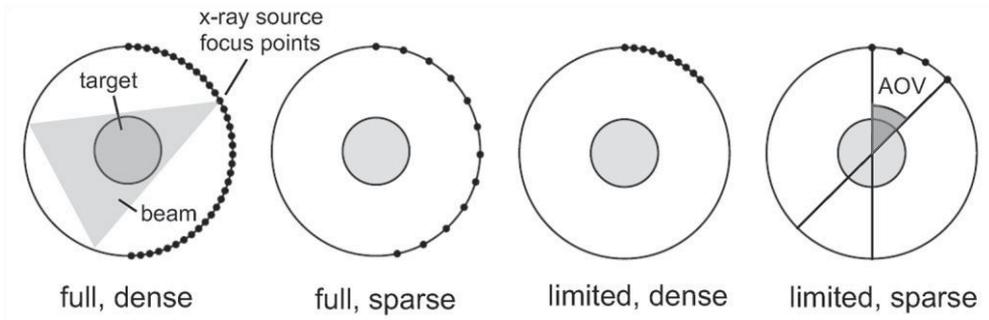


Figure 8. NVBT projections are collected in limited dense or sparse configuration (for Publication II in 11 scans of 4-5 degree spacing. The angle of view (AOV) marks the range of angles where the NVBT is theoretically able to reconstruct edges.

The lack of information appears in NBVT images as false edges and blurring (Figure 9). These false edges or other features that appear images and do not exist in the real target are commonly called image as artefacts.

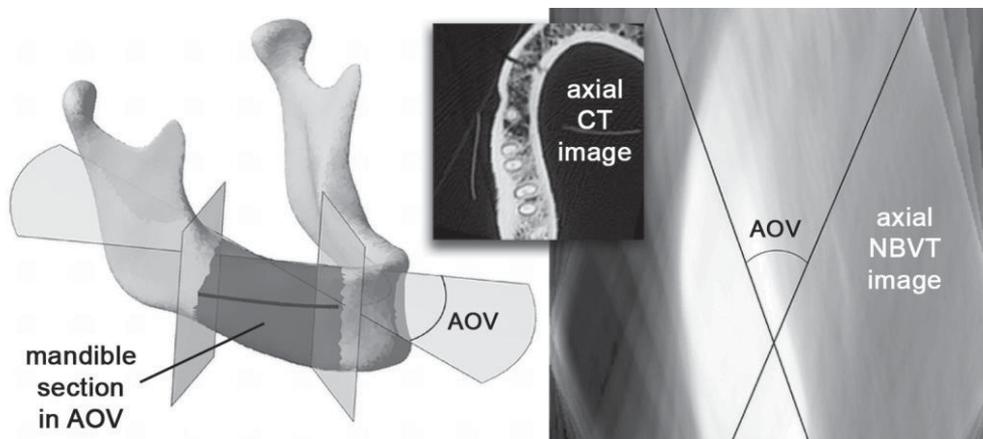


Figure 9. NBVT image reconstruction from limited information. A mandible section that has edges with their tangential orientation in AOV (left) can be reconstructed from an image. Due to the limited angles in NBVT (right), angles outside the AOV are difficult or impossible to reconstruct. Limited information causes ghosting and false edges. An axial CT image (center) reconstructed with full information is provided for comparison. Modified from Publication II. © IEEE 2007.

2.2.3 Cone-Beam Computed Tomography

CBCT was developed as a lower-cost, lower-dose alternative to conventional CT (Worthington et al., 2010). In CBCT a C-arm equipped with a rectangular image sensor and a x-ray source forming a cone-shaped beam performs a single and fast half-rotation around the target. This setup differs from conventional CT (sometimes called fan-beam CT or multi-slice CT) where a strip-like sensor completes several rotations around the target in a helical pattern. CBCT is a good alternative for CT in applications where a high exposure is not desirable and the investment and infrastructure cost needs to be feasible for small private clinics such as dental practitioners. Consequently, CBCT has seen wide use in dental and maxillofacial imaging.

The small exposure of CBCT and the use of cone-shaped beam geometry leads to some tradeoffs in image quality (Liang et al., 2010). Lower exposure leads to lower signal-to-noise ratio. The cone-shaped beam and the larger detector make the image more susceptible to scattering and beam hardening artefacts. Both CT and CBCT are especially prone to metal artefacts (Vannier et al., 1997), (Us, 2019) that are very common and a major factor in reduced image quality in dental applications (Figure 10).

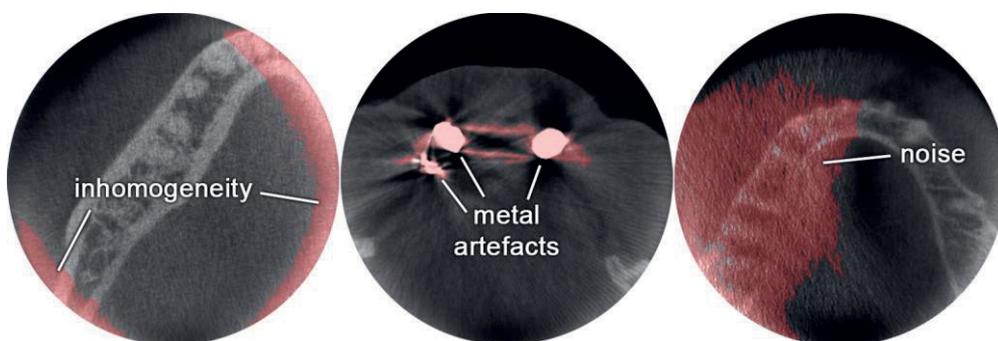


Figure 10. Examples of artefacts in CBCT images. Areas of inhomogeneity (left), metal (center) and noise (right) are highlighted in red. Modified from Publication IV.

Metal artefact reduction (MAR) (Us, 2019) can be applied to mitigate the effects of streaking and ghosting caused by fillings, braces and other metallic or hard objects that have poor x-ray penetration.

CBCT typically uses algebraic image reconstruction techniques (ART) (Andersen, 1989) instead of the more traditional filtered backprojection (FBP) that is still common in CT. Since CBCT image intensities are not directly computed on x-ray

attenuation the gray-value mappings to generally-used Hounsfield Units (HUs) may not be straightforward. Known HU values are valuable for example in CT-based bone density estimation (Schreiber et al., 2011) and are useful in simple image segmentation such as thresholding. The lack of reliable HU scaling limits the use of simple segmentation techniques such as threshold for CBCT.

2.2.4 MR Imaging with High-Intensity Focused Ultrasound

MRI is a versatile modality used in a variety of structural and functional imaging tasks (McRobbie et al., 2006). The physics of MR imaging are based on the phenomenon of Nuclear Magnetic Resonance (NMR) where the atomic nuclei, most commonly hydrogen are excited by microwaves to have them emit back weak, but detectable signal. In MRI no ionizing radiation is needed and a good contrast between a range of soft tissues can be achieved.

Two special properties make MRI a good modality support to HIFU therapy. The heat accumulation inside the body can be monitored through the shift in the resonance frequency of the protons in the water (Köhler et al., 2009) and the blood perfusion to the tumor can be monitored with Gadolinium injection (Burn et al., 2000). A goal of this work was to segment both parts of the tumor, the (Gadolinium-) perfused volume (PV) and non-perfused volume (NPV) that represent the untreated and the treated part of the tumor (Figure 11).

The large ROIs that are often needed in HIFU can take time to acquire. The images may suffer from patient movement artefacts due to the discomfort of receiving the HIFU therapy. Like CT and CBCT, MRI is also prone to metal and inhomogeneity artefacts, although these will appear in MRI differently and thus require modality-specific pre-processing.

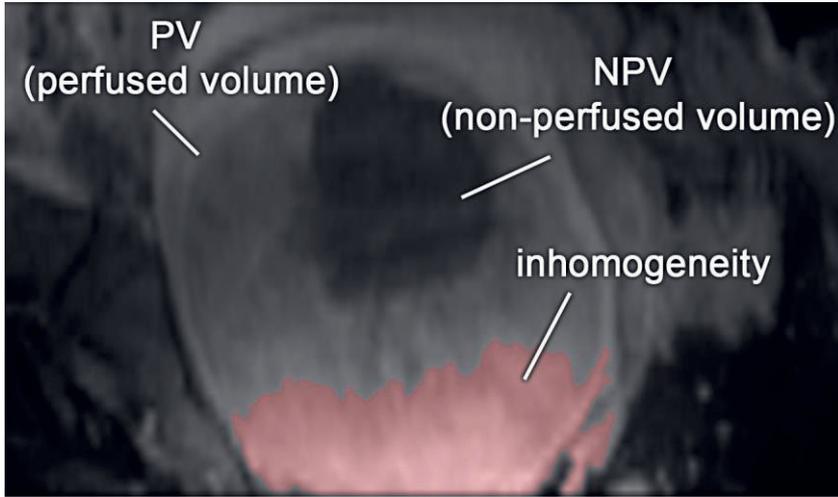


Figure 11. Post-operative MR-HIFU image of a uterine fibroid. The (Gadolinium-) perfused volume will show in light gray, and the non-perfused volume in dark gray. An MRI inhomogeneity artefact is highlighted in red. Modified from Publication V.

2.3 Error metrics for a segmentation algorithm

The following chapters will discuss the performance of the presented segmentation algorithms against some reference, usually surfaces or volumes drawn by a human expert. In case of a binary segmentation, where all voxels are labeled representing either a target or background (unlabeled), probably the most commonly used metric is the Dice similarity coefficient (DSC) (Dice, 1945), (Sørensen, 1948)

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (1)$$

where X and Y are the sets of voxels labeled as the target in the segmented and reference images.

For surface meshes that have point-to-point correspondence, a simple and popular metric is the sum of squares (SS) (Arun et al., 1987)

$$SS = \frac{1}{N} \sum_{i=1}^N \| \mathbf{p}'_i - \mathbf{p}_i \|_2 \quad (2)$$

where \mathbf{p}'_i and \mathbf{p}_i are the coordinate vectors of the nodes in the evaluated and reference meshes that both have N nodes. $\| \cdot \|_2$ represents the Euclidean norm. SS

may not be practical in cases where the segmentation does not fit a pre-constructed surface model and maintain point correspondence through the dataset. For comparing surfaces without point to point correspondence the mean surface distance (MSD)

$$d(\mathbf{p}, S') = \min_{\mathbf{p}' \in S'} \|\mathbf{p}' - \mathbf{p}\|_2 \quad (3)$$

$$MSD = \frac{1}{N_S + N_{S'}} (\sum^{N_S} d(\mathbf{p}, S') + \sum^{N_{S'}} d(\mathbf{p}', S)) \quad (4)$$

or some variant is commonly used. MSD is the standard error metric all publications of this work. There S and S' are the sets of nodes belonging to the evaluated and reference surface meshes. Another popular metric is the root mean squared (RMS) distance

$$RMS = \sqrt{\frac{1}{N_S + N_{S'}} (\sum^{N_S} d(\mathbf{p}, S')^2 + \sum^{N_{S'}} d(\mathbf{p}', S)^2)} \quad (5)$$

Most articles related to this work present their results either with MSD (exact definitions differ slightly) or RMS. Hausdorff distance is yet another popular surface error metric in medical image analysis although less used the in the literature of dental image processing.

2.4 Prior work

Image segmentation is a vast research topic. A Google Scholar search (as of 23th of Feb 2020) with terms “medical image segmentation” yields 1,34 million results. To outline the scope to dental segmentation, only methods that target facial bones either mandible, maxilla or both are discussed in the following sections. The scope is further limited to works that report quantitative results in some metric and have CT or preferably CBCT as the modality. A few exceptions of this in studies that bring clear methodological value are included. Some of the methods presented here take bone segmentation as pre-requisite for other purposes such as landmark detection, tooth or mandibular canal segmentation. These studies offer comparisons to Publications I to IV.

A brief review of literature for methods of segmenting the uterus in MR images will also be conducted. This part is less comprehensive since it has less weight in the

results of this work and simply because there is rather little published research available for the topic. These works discuss problems similar to what is presented in Publication V.

2.4.1 Facial bone segmentation

In short, the research of dental CT and CBCT segmentation was first and for rather long dominated by methods based on or inspired by active shape models (ASMs) (Cootes et al., 1995), other statistical shape models (SSMs) and active appearance models (AAMs) (Cootes et al., 2001). This was followed by a period when Multi-atlas (MA) methods (Aljabar et al., 2009), (BachCuadra et al., 2015) were dominant. Results with convolutional neural networks (CNNs) such as the U-net (Ronneberger et al., 2015) and other deep neural network topologies have been presented only fairly recently. Level sets (Malladi et al., 1995) and path following algorithms (Dijkstra) have been applied over time for mandibular substructures (teeth, mandibular canal). Local and global thresholding with threshold estimation methods such as (Otsu, 1979), and variants of Gaussian mixture models (GMMs) such as expectation maximization (EM) (Moon, 1996) have been proposed both early and lately.

Early research with statistical shape models

An early work by (Vannier et al., 1997) discussed the use of CT in dental and maxillofacial imaging and the factors affecting its usability for tasks such as surgical and implant planning. The image quality factors including metal artefacts and the chosen reconstruction technique were debated in this work also in the context of image segmentation.

(Miet Loubele et al., 2006) were one of the first to present a method and report results for dental CBCT segmentation. They used global thresholding and determined the thresholds heuristically with histogram analysis (Baillard et al., 2001). The CBCT segmentation results were validated against CT and small, but significant differences in bone thickness across the mandible, maxilla and substructures were reported. Around the same time (Rueda et al., 2006) presented a novel AAM that was trained with a sizeable set ($n = 215$) of CT images to segment the cortical bone, trabecular bone and the mandibular nerve. During the same the same year (Lamecker et al., 2006) presented an SSM for mandibular segmentation in CBCT. The handling

of the surface model in this paper is rather similar to Publications I-III. This work introduced a mandible surface model with separated substructure regions, which enabled the model to omit teeth due to the special challenges they posed (left/right, teeth, condyles). They were able to report a consistent sub 0.35 mm mean distance error from segmented mandibular regions to an expert drawn surface.

(Kainmueller et al., 2009) upgraded the SSM presented in (Lamecker et al., 2006), integrated the mandibular nerve canal to the surface model and trained the algorithm with a set of CBCT images ($n = 109$). A Dijkstra-type algorithm (Stein et al., 1998) was used to track the canal. This work was able to match the accuracy of (Lamecker et al., 2006) and exceed (Rueda et al., 2006) in tracking the mandibular canal. (Kainmueller et al., 2009) remains one of the most referenced early work on in mandibular segmentation.

Together (Lamecker et al., 2006), (Rueda et al., 2006) and (Kainmueller et al., 2009) form the most relevant comparison to Publication I-III due to the similarities in the problem setting and used methodology.

As an interesting and notable departures from the SSM/AAM -based algorithms for the period are (Le et al., 2009) that presented a graph cut (Boykov & Funka-Lea, 2006) based, interactive segmentation technique for both mandibular and maxillary regions including the teeth, (Barandiaran et al., 2009) who proposed a mandibular segmentation pipeline that consisted of MAR filtering and (Reddy & Kumaravel, 2010) who took an unique approach with gray-level co-occurrence matrices and curvelet transform for voxel classification.

(Gollmer & Buzug, 2012) presented another SSM with unique shape priors and concluded that the determinant of covariant as proposed in (Kotcheff & Taylor, 1998) performed the best. Most notably, they were able to reach 0.50 mm surface error (including mandible but no teeth) with a rather small shape atlas ($n = 30$). (Duy et al., 2012) continued the work in (Kainmueller et al., 2009) by presenting an SSM method for individual tooth segmentation and the earliest method found in this review capable of segmenting the maxilla in CBCT.

(Lloréns et al., 2012) presented a method developed for the specific purpose of implant planning. The method first estimated the placement and curvature of the dental arc and generated both orthopantomographic and cross-sectional views of the mandible. The goal and approach taken in this work is very similar to Publication IV. These views were further used to reconstruct the mandible (by thresholding), the mandibular nerve (by fuzzy connectedness ((Udupa & Samarasekera, 1996)), cortical bone and the trabecular bone. Their work is distinct for its use of the MPR-type

space (Preda et al., 1997) curved around the mandibular arc instead of the usual Euclidean voxel grid.

(Chang et al., 2013) presented a method for extracting an anterior wall segment of the maxilla in high resolution. This work introduced a wavelet density model (WDM) for generating the shape descriptors in an SSM.

Multi-atlas segmentation

An article by (Wang et al., 2013) introduced the multi atlas segmentation methods for dental imaging. Their core idea was to create the atlas (Aljabar et al., 2009), include local context beyond single voxel comparisons by grouping the voxels in patches (Coupé et al., 2011), (Rousseau et al., 2011) and use maximum a posteriori (MAP) estimation for assigning the voxel labels. The authors also introduced a strategy for integrating CT (n=30) and CBCT (n=15) based atlases. Wang et al. improved their method further (Wang et al., 2014) by introducing landmark-guided registration for better anatomical correspondence in the atlas generation. The work also discussed the effect of the atlas size on the segmentation accuracy. (Wang et al., 2014) was one of the first to include teeth and parts of maxilla were in the segmentation. The downside of many atlas-based based methods was their slow computation time, in this work approximately 5 hours per volume. The authors cite the need to improve the computation time as one of their motivations in (Wang et al., 2016) to develop an RF-based (Breiman, 2001) classifier. They extracted a number of local and non-local, Haar-like (Viola & Jones, 2001) features based off their training atlases. With the RF-based, local segmentation they were able to both cut computation time to 20 min and improve the segmentation accuracy.

(Abdolali et al., 2017) proposed yet another SSM for the mandible and the mandibular nerve canal. They found that conditional SSMs (Yokota et al., 2013) improve segmentation accuracy over conventional SSMs. An edge-preserving noise suppression technique of low-rank decompositions (Ong & Lustig, 2016) was also used in pre-processing.

(Fan et al., 2018) presented a straightforward method where they created a single image-template for the mandibular bone with a number of markers for the bone and background, performed a non-rigid registration with the FNIRT algorithm of the FSL toolbox (Jenkinson et al., 2012) to propagate the labels of the template to new samples. (Chuang et al., 2018) revisited the idea multi-atlas segmentation implemented in diffeomorphic, FSL-powered registration. Their work was the first to use N4 bias field correction (Tustison et al., 2010) to preprocess mandibular CTs.

The authors reported remarkable segmentation accuracy, but stated also that the human assessed quality of automatic segmentations still clearly trailed the manual segmentation in all tested substructures.

(Cuadros Linares et al., 2019) presented a concept of super-voxels, regions of raw voxels locally grouped together based on a similarity measure and by k-means clustering. They then construct a weighted graph on the relations of these super-voxels. The super-voxels were meant to reduce the size of the graph and thus computational complexity. The binary segmentation part was based on graph-cuts initialized with minimum-effort manual seed inputs (a few brush strokes on the bone and background).

(Indraswari et al., 2019) presented a segmentation pipeline where they first perform coarse segmentation for the bone with thresholding and then fit a polynomial to gain an estimate of the curvature of the dental arc that they use for the eventual reconstruction of 3D surface. For the thresholding they compared a GMM against Otsu's and hierarchical cluster analysis (Arifin & Asano, 2006) and found GMM to perform best. The authors also compared the results to the U-net and found it accurate but underestimating the size of the bone.

(Vaitiekūnas et al., 2019) proposed local thresholding in a sliding window using Otsu's with spatial filtering, reconstructed the bone surfaces with run-length encoding (Curless & Levoy, 1996) and benchmarked the result against manual segmentation in pre- and post-operative CBCT sets with remarkable accuracy of 0.057 mm MSD.

The emergence of deep learning

(Zhang et al., 2017) were the first to test the U-net (Ronneberger et al., 2015) and CNNs to a reasonably-sized dataset (CT $n = 107$, CBCT $n = 77$). This work followed (Wang et al., 2014) and again combined bone tissue segmentation with anatomical landmark detection. The pipeline consist of two CNNs in series. The first constructs distance maps to chart the landmark placement configuration and the second uses the distance maps and the image input to generate bone segmentation maps and landmark heatmaps. A major improvement in both mandibular and midface segmentation over their previous multi-atlas (Wang et al., 2014) and RF (Wang et al., 2016) was reported in the paper. The study also concluded that it is beneficial to solve the segmentation problem and landmark detection problem simultaneously. The authors reported identical results for the bone segmentation later in (Zhang et al., 2020). This work presented a more thorough validation and strengthened the

earlier conclusion on the benefits of solving the segmentation and detection problems simultaneously.

Also (Torosdagli et al., 2019) chose to combine mandibular bone segmentation and landmark detection. This approach differs from (Zhang et al., 2017) in that it solves the segmentation problem first and passes the output for the landmark detection. For the segmentation the authors chose a densely connected and very deep, 19-layer CNN (Jegou et al., 2017). After segmentation, the pipeline used U-net to generate landmark probability maps and LSTMs (Hochreiter & Schmidhuber, 1997) to localize the landmarks. The authors reported a slight advantage of the 19-layer CNN over the U-net used in (Zhang et al., 2017) in two independent datasets and statistically significant similarity to manual reference segmentation.

(Minnema et al., 2019) validated four segmentation methods on a CBCT dataset corrupted with metal artefacts. Three of these were CNNs: an MS-D (Pelt & Sethian, 2017), a modified U-net and a ResNet (He et al., 2016) and the last one was an ACM (Snake evolution) (Yushkevich et al., 2006). The three CNNs were equally accurate and beat the Snake evolution method with a clear margin. The MS-D only used about 46 thousand trainable parameters vs. 15 million in U-net vs. the 37 million parameters in ResNet.

2.4.2 Image view synthetization

A dental radiologist or an implant planner needs to have the correct views on the areas to be operated on the pre-operative CBCT images. A typical set of commonly used views include a (primary) orthopantomographic view over the dentistry (Figure 6) and a number of (secondary) cross-sectional slices over the target area (Figure 2). A single CT or CBCT acquisition can be used to synthesize a countless number of views but this process can take a considerable time of drawing, re-orienting and configuring. A segmentation method can automatically fit or assist in fitting a frame of reference according to which the desired views should be generated to patient anatomy and pose. In order to be practical the segmentation method should be fast enough to not keep the operator waiting for extended periods during the planning workflow.

One of the first complete image synthetization pipelines including image processing, shape modeling and multi-view generation was presented in (Cucchiara et al., 2004). The authors formulated a thorough mathematical framework for detecting surgical stents in images and discuss the measurement error related to

implant planning whether the views are re-oriented according individual anatomy or not. (Tohnak et al., 2006) took a rather unique approach where they simulated the actual panoramic imaging geometry including the reconstruction. (Chanwimaluang et al., 2008) discussed the detection of the dental arch through local thresholding and morphological operations. This article presented an alternative solution to a problem that is almost identical to our Publication III. (Akhoondali et al., 2009) constructed another synthetization pipeline methodology very similar to (Chanwimaluang et al., 2008). Their work also discusses the emulation of x-ray attenuation parameters to gain an appearance that better matches a native scan. (Han et al., 2011) extended the work of (Tohnak et al., 2006) by refining the pipeline by automating the image processing operations.

(Sa-Ing et al., 2013) were the first to present a large ($n = 120$) dataset for panoramic image synthetization. Their article recycles much of the methodology (thresholding and morphological operations) proposed since (Cucchiara et al., 2004). CBCT image synthetization is further discussed by (Luo et al., 2016). Their described application is a very close match to ours in Publications III-IV. (Papakosta et al., 2017) presented a proper quantitative learning and validation strategy for modelling dental curvature. Again the same rudimentary segmentation (thresholding and morphological operations) was exploited in (Yun et al., 2019). The main novelty of this work is in how the thickness of the dental arc was estimated and how the thickness parameter was implemented for the synthetization. The work (Bae et al., 2019) puts special emphasis on fitting the occlusal plane (a plane that fits to gap between upper and lower teeth when the mouth is closed) and introduces a B-spline-based modelling approach for the dental arc estimation. The spline modelling aspect presented the paper is similar to what we used in Publication IV. It differs from ours in that the spline was mostly used for smoothing the segmentation result where ours is also for filling gaps and holes and in the type of spline selected (B-spline vs. thin plate).

2.4.3 Muscle tumor segmentation

Pre- or post-operative MR-HIFU fibroid segmentation is a sparsely studied topic. To gain wider understanding of the background, also studies of the segmentation of a healthy uterus are covered in the following section. After an early work by (Yao et al., 2006) only two known distinct tracks of research on the topic, the first started in (Fallahi et al., 2010) and the second in (Militello et al., 2013) have been conducted.

The methods proposed in (Fallahi et al., 2010) in (Fallahi et al., 2011) are based on an evolution of the fuzzy c-means (FCM) algorithm (Pal et al., 2005). They first acquire a rough tissue classification by the non-probabilistic version of the FCM, refine the result with morphological operations and then apply the modified version of the probabilistic FCM for the final results.

The range of methods proposed in (Militello et al., 2013), (Militello et al., 2015) and (Rundo et al., 2016) evolved from region growing with semi-automatic seed points through FCM combined with global thresholding and finally back to region growing with automatic seed selection by a divide-and-conquer type of search. The same authors have also put out two review papers (Militello et al., 2014) and (Rundo et al., 2019).

Both (Yao et al., 2006) and (Khotanlou et al., 2014) discuss and propose the use of level sets in the fibroid segmentation. (Yao et al., 2006) starts from a manually defined seed points, applies fast marching level sets and refines the result with Laplacian level sets. The work by (Khotanlou et al., 2014) is notable for the use of ACM-style shape priors. (Kurata et al., 2019) proposed the only known deep learning approach by validating a U-Net on a large ($n = 122$) set of fibroids.

3 MATERIALS AND METHODS

3.1 Materials

The data collected for this work is presented in Table 1. Our work began with small sets of images scanned with the still evolving NBVT modality. The work continued with two sets of prototype CBCT images acquired from the manufacturer and a dental clinic. The MR-HIFU images were collected from four clinical sites participating in the development of the uterine fibroid therapy.

Table 1. The datasets of this work summarized.

Pub.	n	modality	notes
I	3	NBVT	Initial dataset with 3 patients. A surface model for the mandible was built using 31 independent MRI images.
II	9 + 4	NBVT	A follow-up dataset with 9 dry skulls and 4 patients.
III	14	CBCT	The first CBCT dataset.
IV	19+30	CBCT	Two sets of CBCTs with prototype scanners of two different generations, a set from each.
V	29	MR-HIFU	Data collected from 4 sites with a single equipment manufacturer.

3.2 Image processing pipeline

Solutions to image analysis problems usually consist of more than a single operation or an algorithm. The algorithm(s) executing the segmentation usually benefit from some sort of pre-processing of the raw image input. The primary motivation for pre-

processing is in making the data set more uniform. Common pre-processing steps include resampling the image to common voxel size, normalization of the intensity values according to a common reference and correcting the size, location and orientation of the target by affine co-registration. These are often useful operations regardless of the chosen segmentation method. The chosen combination of operations are typically configured to a series of consecutive steps where one operation takes the input from the previous one. This series is commonly as called the image processing pipeline. Figure 12 presents an example composite of a pipeline that presents a summary key components and task used in this work.

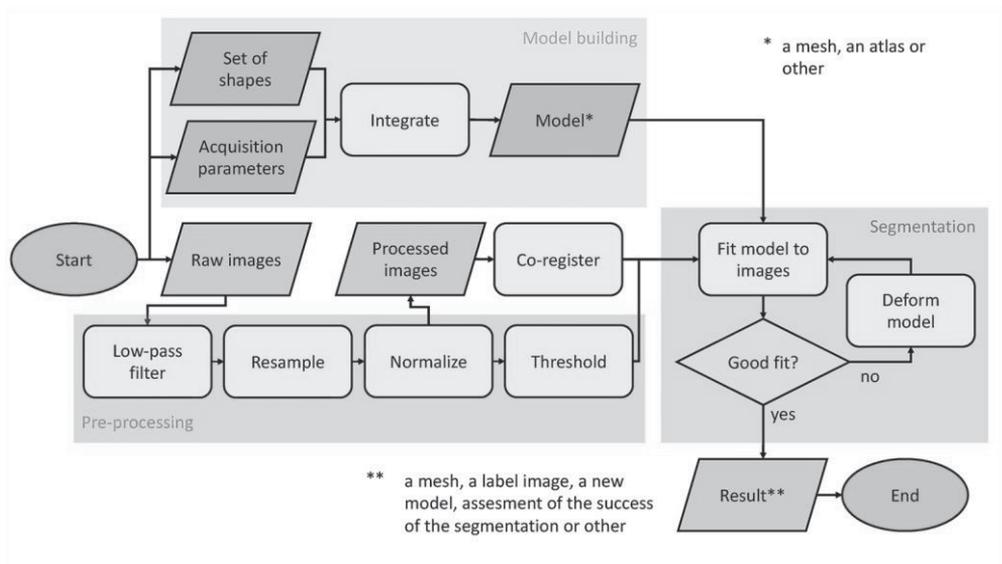


Figure 12. An image processing pipeline. This is an example of a pipeline put together of the typical components implemented in this work.

The following sections present an outline of the development of three major versions of the image processing pipeline built for 1) NBVT (mostly Publication I, also parts of II and III), 2) CBCT (Publication IV) and MRI (Publication V). These are generalizations since all publications present their own solution on the current problem. The following sections will present the evolution pipelines during the accumulation of more input data and the updated application requirements.

3.3 Model building

Image segmentation can be defined as a problem where a (3D) surface or (2D) contour will be fitted on a target. All pipelines implemented in this work include at least a subproblem that was formulated this way. The prior information brought to the segmentation outside from the current image is generally called as the model. In practice the term model can mean various things. In a basic form the model can be a single surface drawn on one input image to be fitted on the rest in the dataset. An important motivation to build a model is the opportunity to include some characteristics of the target shape such as a typical (average) shape (Publications I-III), the variation of the shape around it and then to penalize the segmentation algorithm from trying to conform the surface to abnormal or non-natural shapes (Publications IV-V). Some of the influential and early work on fitting contours was published (Kass et al., 1988) in ACMs and on the incorporation of shape information by (Cootes et al., 1995) in ASMs. A model can also incorporate elements that are not part of the input image, such as some frame of reference or a coordinate system (Publications I-III).

Our model building work for the NBVT pipeline (Publication I) began by taking an independent set of 31 MRI images of the whole head and drawing the mandible manually on all images. This yielded one binary label map per image where a voxel was given either label 1 (mandible) or 0 (other). The MRI image set were then co-registered with a non-rigid deformation algorithm (Rueckert et al., 2003), the deformations applied to the label maps and the maps converted to triangular mesh surfaces (Lorensen & Cline, 1987). From this set of shapes (meshes) an average surface for the mandible created by selecting one mesh as the template and co-registering the rest on the template before taking the average. This first iteration of the model included also an average estimate of the mandibular centerline. The centerline was to be deformed along the surface and to provide a simple coordinate system where orthogonal slices would later be fitted for visualization.

The first model was rebuilt and re-aligned using (Umeyama, 1991) from the original 31 MRI-based mandibles and a fresh set of mandibles drawn on 9 independent CT images (Publication II). The updated model now included the complete mandible including the condyles (the joints along with the mandible pivots) and was of higher resolution (7407 nodes, 14810 triangles). A set of 5 round fiducial markers representing the radiopaque landmark pellets that were embedded in the bite plate held by the subject in the mouth during NBVT scanning (and thus visible in the images) was added to the model. The pellets were used to initialize the model

by placing it roughly on the target location and orientation. The new model introduced also a division of fixed surface areas of exterior, interior mandible and the teeth and condyles (Figure 13). The division to areas was added as we foresaw the need of treating these areas differently during segmentation due to anatomical variation and varying effect of image artefacts (mostly metal) on different areas.

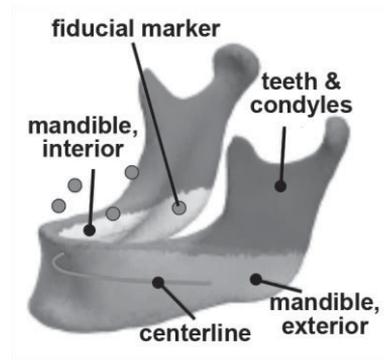


Figure 13. An average surface model for mandible. Modified from Publication II. © IEEE 2007.

The image modality was eventually changed from NBVT to CBCT but the target for the segmentation remained the same (Publication III). All components except for the 5 markers only visible in the newest NBVT images of the prior models were retained. The mandible surface model was slightly upgraded by adding the surface tangents vectors of the mandible exterior and interior. This was done to test if the surface tangents could be used as estimates of the local teeth root orientations. The estimates were added to aid visualization and not to be used in the segmentation. With this model version we also experimented the use of mandible PCA-based shape variation modes according to (Cootes et al., 1995), but we did not end up using them. Our decision was based on the assessment that although the modes may have improved the results slightly, they add complexity and increase the computational time.

More CBCT images were added to the data set and the segmentation target was updated to include also the maxilla and potentially the maxillary sinuses (Publication IV). The inclusion of maxilla added more target area and most importantly, more shape variation. The image co-registration methods used for the mandible in Publications I and II did not yield satisfactory results after the maxilla was included. We gained some success by using the multi-atlas method by (Lötjönen et al., 2010) but it was deemed too slow even though this it decreased computation times dramatically from previously available methods. Due the lack of satisfactory results

with multi-atlas segmentation and especially image co-registration, we ended up abandoning the approach of using a pre-built surface model. Instead we chose to build a data-driven model solely based on the characteristics of the problem, some general assumptions of the target and its orientation in the image.

Our updated goal (Publication IV) now was to segment the exterior of all visible facial bones (maxilla, mandible, teeth and zygomatic bones). To accomplish this, we decided to try an approach of fitting a mesh that is generated during the segmentation. To implement this we devised an approach where potential mesh points one soft-tissue to bone edge are searched along a set line profiles. The use of line profiles provided us performance gain since only the image voxels on the profiles and not in the whole volume need to be evaluated. These line profiles are organized to a grid that is fitted (bent) along the contours of the face (Figure 14).

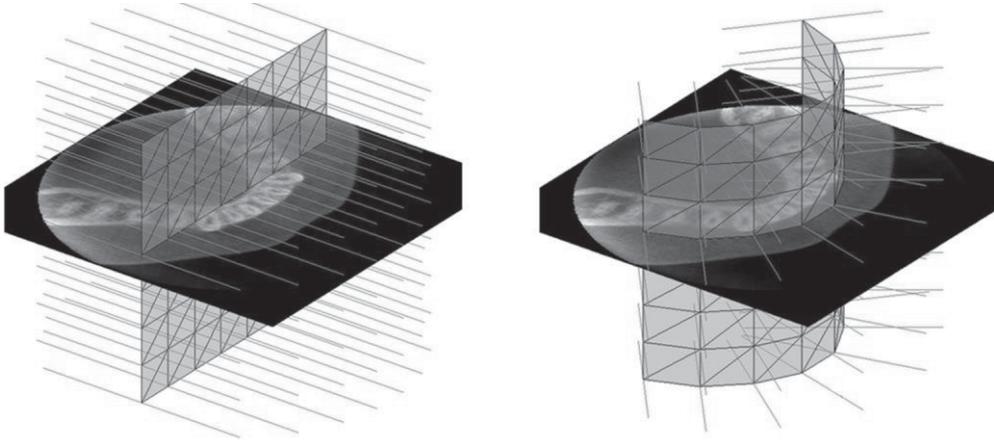


Figure 14. The procedure for fitting a set of line profiles on the facial bones. On the left is the grid in the initialized position (a plane splitting at the center of the volume) for the first iteration. On the right is the grid after bending along the contours of the face.

The process of fitting and re-fitting the grid and the line profiles is called as surface parameterization (Publication).

After developing the CBCT segmentation method our focus was shifted to MR-HIFU and the segmentation target was set to uterine fibroids (Publication V). Despite of these differences we decided to test and then to revise the data-driven approach developed for CBCT (Publication IV). The change of the target required us to switch the surface parameterization topology from an open surface (the bent plane) to a closed one (a sphere-like). The fibroid segmentation problem forced us to rethink the model initialization. Using an analogy to similar the facial bones

segmentation we could have initialized the new model by fitting a sphere to the approximate location of the fibroid. However, in the MR-HIFU problem we had an extra advantage in the ultrasound sonication treatment plan. This plan is a mandatory step in order to perform the treatment. This plan includes a cluster of three-dimensional coordinates of the focal points (sonication cells) where the ultrasound sonication is to be administered. What is even better, the plan includes the shape parameters of the sonication cells including size and orientation. This enabled us to construct an initial model that is a representation, even though coarse, of the treatment volume and thus much of the tumor itself. We created the initial model surface from the sonication cells by first rendering them as 3D meshes, wrapping a convex hull on the point group (Barber et al., 1996) and inflating the convex hull to match the assumed safety margin (10 mm) from the cells to the edge of the tumor (Figure 15).

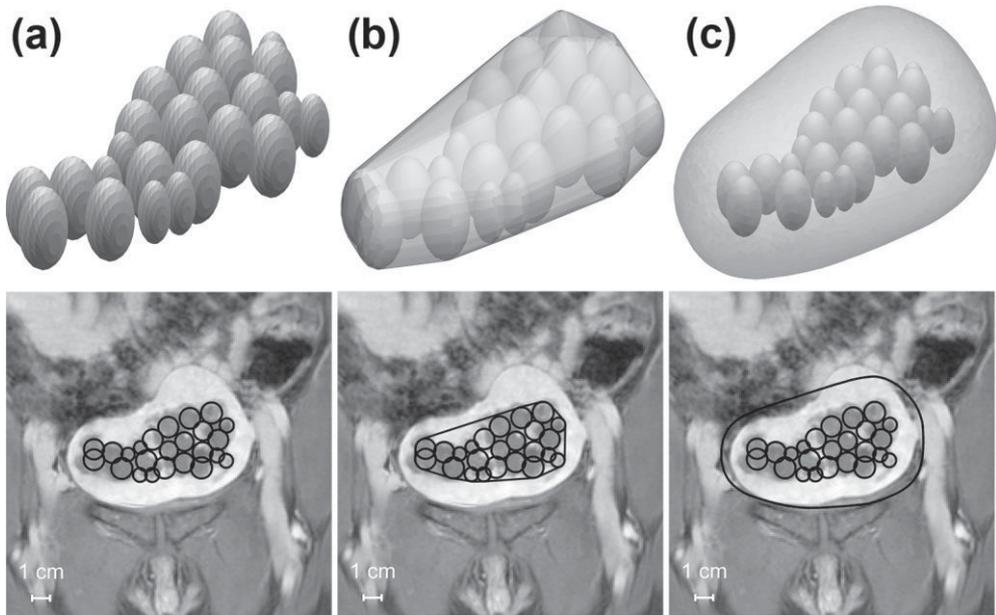


Figure 15. Building of a surface model for the PV segmentation. The model is based on sonication cells (a) extracted from the therapy plan. A convex hull is wrapped around the cells (b) and expanded (c) according to the safety margins in the treatment protocol. Adapted from Publication V.

Similar to Publication IV, this surface was used for the parameterization and not to segment the surface.

3.4 Image pre-processing

The purpose of image pre-processing is to make the set of input images uniform and clean by eliminating or normalizing factors that are rather a characteristic of the modality, result of the scanning setup or random error and not a property of the target. Typical pre-processing steps of this work included resampling to standard resolution, suppression of noise and other common image artefacts and normalization (equalization) of the gray value distribution and correction of image inhomogeneity.

Selection of the best pre-preprocessing steps usually requires designer expertise and may involve heuristics. The selection process is often implemented, as in this work, by trial and error by assessing the visual appearance of the result image set and monitoring the segmentation algorithm performance and robustness. This step may need to be revisited once the properties of the image data are known and the most promising candidates for the actual segmentation algorithm have been chosen.

3.4.1 Image filtering

Images will practically always contain a component of noise of varying strength. In medical imaging the presence of noise usually is due to the tradeoff between acquisition time and image quality. In CBCT short acquisitions mean smaller exposure to the ionizing x-ray radiation. In MRI fast scan sequences help in managing the effect of natural body movement (heartbeat, breathing) on the image quality and ease the discomfort of lying in a loud, tightly confined space. These are examples of constraints that will set the practical limits on how much information can be collected during scanning and thus the amount of noise that need to be tolerated during the image reconstruction and pre-processing. Image noise is a typical consequence of this uncertainty.

In Publications IV and V we filtered the images with a Gaussian kernel (Haddad & Akansu, 1991) to dampen the noise component in CBCT and MRI. The main benefit doing of this is to mitigate the effect of noise on the calculations to find the edges during the segmentation. Approaches like nonlocal means (Buades et al., 2005) could provide better edge preservation over Gaussian kernels with effective noise suppression. However, these were not used in this work because of the their higher computational demand limited benefit over the simpler Gaussian method.

The other motivation to use Gaussian filtering in Publication V was to reduce potential aliasing prior to image resampling. Aliasing errors are not very common when resampling images of biological structures. Still, we decided that it is a good practice to lowpass filter the images and choose the lower limit of the kernel size according to the Nyquist-Shannon theorem (Shannon, 1949) prior resampling.

3.4.2 Resampling to a uniform grid

Resampling the image data to a common-sized voxel grid may not be a requirement, but is often done for convenience. An isotropic (equally spaced) voxel grid is useful in image co-registration and surface model fitting tasks. For example, it simplifies the affine transformations (translation, rotation, scaling, shear) to align the model to the target (Publications I to III). Our raw data in the NBVT and CBCT pipelines (Publications I to IV) was reconstructed to an isotropic voxel grid by default and instead of image resampling, the model was simply scaled according to the voxel size. Our MRI data (Publication V) was not isotropic, so resampling with trilinear interpolation was used.

3.4.3 Equalization

Artefacts are distortions that appear in the image even though they do not exist in the real target. Image artefacts can be a result of the image reconstruction, induced by interfering objects somewhere in the scanner FOV or be a property of the scanning setup. Reconstruction artefacts can appear due to global lack of information such as not having the target scanned from all orientations (NBVT, in Publications I and II), local lack of information due to dense objects such as metal completely blocking the x-rays from reaching the detector and thus information getting lost (CBCT, Publications III and IV) or by x-ray beam hardening and scatter (CBCT, Publications III and IV). In MRI, the images may suffer a from non-homogeneous magnetic field in the scanner (Publication V). In the images in this work the reconstruction artefacts appear as ghosting (NBVT), metal artefacts as sharp streaks (NBVT, CBCT) and the inhomogeneity artefact as smooth and low-frequency variations in image intensity (MRI). Some of our CBCT and to a lesser extent the MRI data was probably affected by patient movement.

Our general strategy of dealing with the artefacts was less to filter them out by preprocessing and more to tolerate them in the segmentation. This most prominent

example of this is the patch-based segmentation strategy in Publication IV where the CBCT image data had a high prevalence of metal artefacts (present in 67% of the images). In Publication II we reduced the NBVT ghosting by normalizing with contrast limited adaptive histogram equalization (CLAHE) (Pizer et al., 1990) (Figure 16).

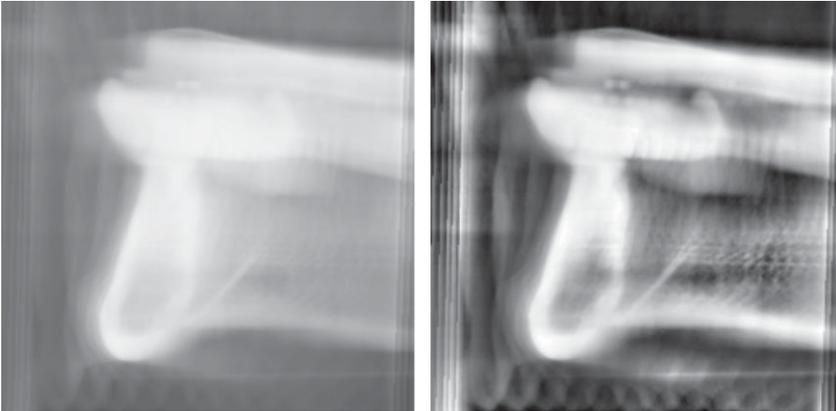


Figure 16. Coronal NBVT images before (left) and after (right) contrast enhancing local adaptive histogram equalization in NBVT. From Publication II. © IEEE 2007.

Although not able to remove the ghosting completely, the visual image quality was improved. The inhomogeneity artefacts that appeared in some of the larger MRI image FOVs in Publication V were preprocessed with a bias field inhomogeneity correction (Tustison et al., 2010). This was necessary since the thresholding algorithms tested in the MR-HIFU pipeline would not have worked consistently with this kind of artefact.

3.4.4 Thresholding

One of the simplest ways of segmenting an image is to select one or more thresholds and label the image voxels according to which intensity range as defined by the thresholds they fall. Global thresholding, selecting fixed thresholds for the whole image, rarely works as the only segmentation strategy but can be useful, even effective when used in the right context and combined with suitable type of pre- and post-processing.

We used global thresholding and morphological operations in Publication I to reveal and pre-process the edges (Figure 17) prior to generating distance maps later in the NBVT pipeline.

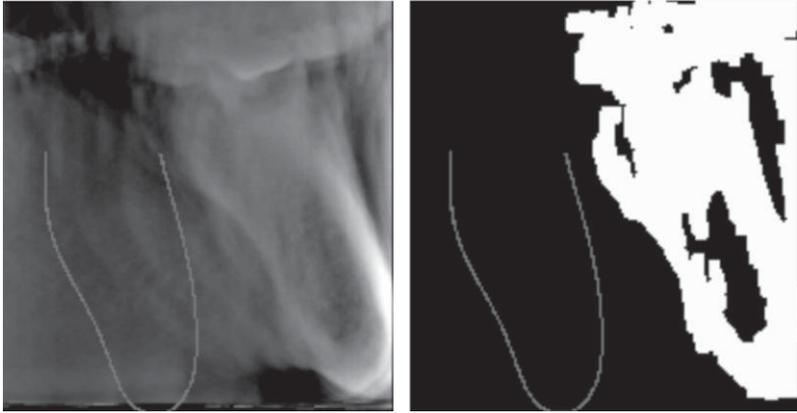


Figure 17. Raw NBVT image (left) is thresholded (right) to reveal the edges in the image. A cross section of the mandibular model (a gray, U-shaped contour) in its starting position prior to transformations is visible in the images. From Publication I. © IEEE 2006.

The initialization of CBCT pipeline in Publication IV needed the approximate intensity ranges of the soft tissue and bone. The thresholds for these were estimated by the direct clustering (DC) method (Pianykh, 2006). The natural division of the CBCT image intensities was to four classes. By this division the assumed classes were (from darkest to brightest) air, soft tissue, bone and tooth enamel combined with metal. Only the ranges for the soft tissue and bone were used. While this method would have not been robust enough to work as a standalone CBCT bone segmentation method, it was good enough to give the starting point to the segmentation algorithm.

3.5 Segmentation

It is often convenient to formulate the image segmentation algorithm as an optimization problem. By doing this the wealth of theories and methods developed for general optimization can be utilized. In very general terms optimization can be defined as a task of finding a minimum of a continuous function given the input and a set of constraints. This function can be interpreted to estimate the energy of a system and the task of the optimizer is to search where the system is at its most

efficient state. For segmentation problems the energy function can be set to describe how well a surface fits on the target and the optimizer is used to find a set of deformations to acquire the best possible fit.

A popular (Kass et al., 1988), (Lötjönen et al., 1999), (Tohka, 2003), high-level definition of the energy function E_{total} , consist two of terms, E_{int} and E_{ext} that are summed for the minimization problem

$$\operatorname{argmin}_{S,I} E_{\text{total}} = \gamma E(S,I)_{\text{ext}} + (1 - \gamma) E(S)_{\text{int}} \quad (6)$$

The external energy component E_{ext} measures the goodness of fit of the surface S of the target in the input image I . An internal energy component E_{int} is usually required to penalize the optimizer from converging to unnatural or unpractical shapes. A weighing term $\gamma \in [0,1]$ can be seen as hyperparameter that can be trained with a set of segmented images. The weigh γ can also be used to prefer certain properties of the output, such as surface smoothness, depending on the end application.

Segmentation problems may contain a large number of parameters, for example coordinates of a node points in a surface mesh. This will easily lead to very high-dimensional spaces unless some strategy to move larger groups of nodes simultaneously is adopted. Publications I to III use a combination of affine (rotation, scaling and translation) that move all model points simultaneously to initialize the model. Skewing, also a degree of freedom of affine transformations, was not used here.

Designing an energy function where there can be a well-defined, guaranteed global minimum would be hard. There are optimizers that are guaranteed to find the global minimum under certain constraints (Kirkpatrick et al., 1983), (Jones et al., 1993), but these are often expensive to solve. The affine initializations used in Publications I to III also help in ensuring that grid search over the limited set remains computationally efficient and a fair approximation of the global minimum can be made. A hyperparameter tuning strategy such as Bayesian optimization could work as well. However, the feasibility such black box function estimators may be limited because the smoothness or the continuity of the underlying energy function may not be guaranteed. The affine initialization forms a good starting point for more flexible, and free-form deformation can be used later to better conform to individual shapes. The results of the segmentation with the affine initialization and after the free-form deformation were presented in Publications I and II separately to help in showing the benefits of using free-form deformations.

3.5.1 NBVT pipeline

Our very first method for the mandibular segmentation was developed with a small set ($n = 3$) of NBVT images (Publication I). This approach followed the principles of ACMs (Kass et al., 1988) and much of the implementation was adapted from (Lötjönen et al., 1999).

The NBVT segmentation pipeline was initialized by first segmenting the bone with a fixed global threshold. The images contained a lot of false edges due to reconstruction artefacts which were suppressed from the thresholded bone label image with two morphological openings. After thresholding, a surface model for the mandible was fitted to the image. The fitting was implemented by constructing six distance maps D of the binary bone label according to (Lötjönen et al., 1999). The distance maps match the six principal orientations of the image volume coordinates (up, down, left, right, back and forward). The values for respective maps are computed by first estimating the three-dimensional orientations of the exterior normals of the edges in the bone labels. For example, the distance map “right” only includes distances to edges whose principal orientation of the exterior normal is to the right. All other five edge orientations are invisible in this map. This way we were able to pull the mesh nodes of the surface model that point primarily to an orientation towards edges in the bone label map have normals primarily in the same orientation (Figure 18).

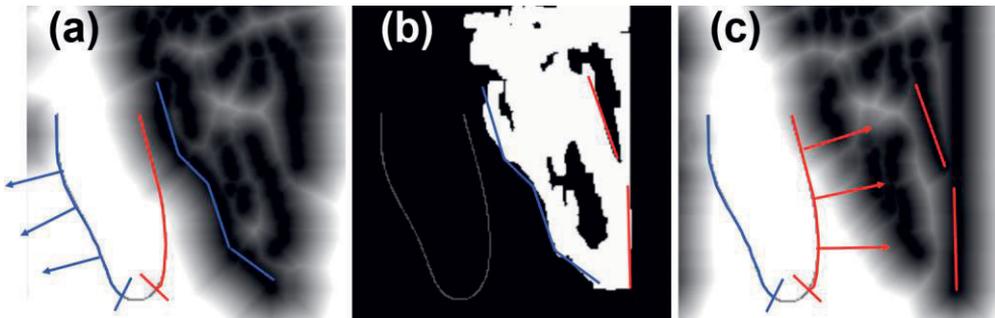


Figure 18. Distance maps from a thresholded image. The distance maps (a,c) are generated from edges of the binary bone label image (b). The maps are generated for all six principal orientations of which the figure shows examples of the “left” (a) and “right” (c) orientations. The left side (in blue) of the surface model is moved towards the edges in left visible in (a) and the right side (in red) to (c). Modified from Publication I. © IEEE 2006.

The maps were then used to make a pre-determined series of affine (rigid) transformations (rotations, translation and scaling) to fit the average surface model (section 3.3) according to

$$E_{\text{ext (affine)}} = \frac{1}{N} \sum_i D_{k(i)}(\mathbf{p}_i) \quad (7)$$

where N is the number of mesh nodes in the model, i the index of the node, $k(i) \in [1,6]$ a function that selects the distance map D that best matches the primary orientation of the exterior normal vector of the node. Once the affine phase was complete there should have been an approximate fit of the model to the target.

The affine transformations are usually too stiff to converge the surface to fine individual shapes. In order to get a better a fit, a series of non-rigid deformations were needed. These were implemented with

$$g_i^{+/-} = -\frac{1}{h} \sum_j^h I(\mathbf{p}_i +/- j\mathbf{n}_i) \quad (8)$$

$$E_{\text{ext (ffd)}} = -\frac{1}{N} \sum_i g_i^- (g_i^- - g_i^+) \quad (9)$$

$$E_{\text{int(ffd)}} = \frac{1}{N} \sum_i^n \left((1 - \mathbf{n}_i \cdot \mathbf{n}_i^*) + \frac{1}{N_{\text{nh}(i)}} \sum_k^{N_{\text{nh}(i)}} \|\mathbf{d}_{i,k} - \mathbf{d}_{i,k}^*\|_2 \right) \quad (10)$$

where $E_{\text{ext(ffd)}}$ approximates the fit of the mesh surface to the original image and $E_{\text{int(ffd)}}$ penalizes from the surface nodes from deviating too far from the normal orientation. \mathbf{n}_i and \mathbf{n}_i^* are the unit surface normals of node i at current and previous iterations, I is the input image, h the surface normal length parameter, $N_{\text{nh}(i)}$ the set of nodes in the neighborhood of node i and $\mathbf{d}_{i,k}$ and $\mathbf{d}_{i,k}^*$ the current and previous displacements of the node vectors. The nodes of the surface model mesh were displaced according a deformable, multiresolution grid (Lötjönen et al., 1999) and iterated to convergence.

A new set of data and slightly updated versions of the energy function components were introduced (Publication II). The affine fitting of the surface model with distance maps (Eq. 7) was dropped and based on fiducial markers (radiopaque pellets) that were detected from the image instead. The marker coordinates were detected by the image reconstruction software independent from the segmentation pipeline. The overall pipeline was kept the same including the use of FFD deformations. After the FFD phase a new elastic smoothing component for the node displacements, according to

$$\mathbf{d}_k = \exp\left(-\frac{\|\mathbf{p}_i - \mathbf{p}_k\|_2}{2\sigma^2}\right) d\mathbf{n}_i \quad (11)$$

was introduced (Lötjönen & Mäkelä, 2001). This updates the displacements \mathbf{d}_k around node \mathbf{p}_k according to displacements d of arbitrary nodes \mathbf{p}_i along their normal \mathbf{n}_i .

Later, this same segmentation method was application for CBCT images (Publication III). Despite the change of the underlying modality, the goal still was to segment the mandible. Eventually only minor tweaks for the pipeline (of Publication II) were required. The affine initialization was re-introduced since there were no fiducial markers in CBCT. Some heuristics was also added in the handling of the energy function output to better cope with metal artefacts. These were more common and increasingly severe in the CBCT dataset. Heuristics also helped in handling the shape variability in the tooth region that was much less visible in the NBVT image sets. Since there is much closer similarity between the segmentation methods in Publications II and III than there is between III and IV, this final evolution of the pipeline is still bundled to the “NBVT” category in the context of this work despite of the underlying imaging modality. The main goal of the Publication III was to demonstrate the use of segmentation in automatic synthetization of panoramic and orthogonal views that are based on individual anatomy and thus not much effort was put in improving the actual quantitative segmentation results. Although the segmentation pipeline did almost completely change by Publication IV, the ideas gained when designing the rendering geometry for the panoramic view synthetization would be a major part on how the facial geometry was parameterized in the CBCT pipeline that was to follow.

3.5.2 CBCT pipeline

A new CBCT dataset ($n = 49$) and a revised goal to segment also the mandible was introduced by Publication IV. The target was now to set to segment all of the exterior of the facial bones (including mandible maxilla and the zygomatic bones) that were visible in the available FOVs. After unsatisfactory results of experiments with extending the NBVT surface model and segmentation pipeline to the maxilla and attempts of applying multi-atlas segmentation a novel, data-driven approach was developed mostly from scratch.

The CBCT pipeline began by filtering the raw images with a Gaussian kernel to suppress noise and pre-segmenting the filtered images with direct clustering (Pianykh, 2006) to get initial, rough estimates of the intensity ranges of target (bone) and surroundings (soft tissue). The CBCT images suffered from a common

inhomogeneity artefact, a wide, smooth, higher intensity (bright) areas on the borders of the cylinder-shaped FOV where bone touches the border. The effect of this artefact was mitigated by removing a thin rim of the outer edge of the reconstruction cylinder prior to the direct clustering.

The novelty of the CBCT pipeline is in the definition of the segmentation problem. The ideas on how the panoramic views were rendered in Publication III were now applied for use in segmentation. When generating the views, the mandible was effectively pulled straight along the mandibular centerline included in the fitted model. The views were then rendered by summing voxels along the line segments aligned according to the normal orientation of the centerline of specified depth. The CBCT pipeline uses this analogy, but the line segments were now exploited to search potential points on bone-soft tissue edge for the eventual surface reconstruction instead of summing pixels for viewing. This allowed us to use an active contour model-like energy function but limit the computational cost by computing the values on only on voxels that cross with the segments. Publication IV calls this process arc length parameterization, but effectively this means bending a plane with a line segments placed equally in a grid on the face (Figure 14). No pre-built model and thus no mandibular centerline are used in the pipeline, so by a slight modification the bend is now fitted along the surface points detected during the previous iteration (Figure 19).

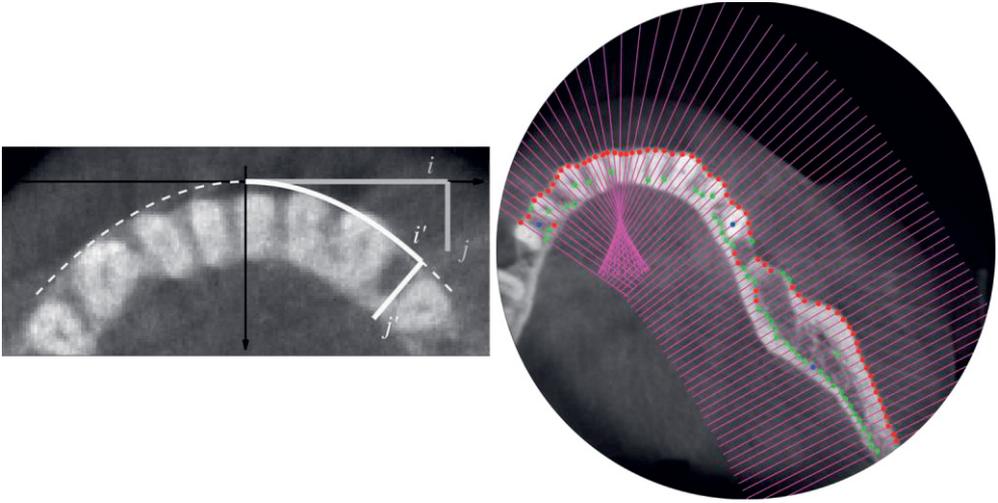


Figure 19. Parameterization of the topology. Parameterization bends a plane (left) along a curved arc. This bending is shown from another perspective in Figure 14. The colored dots (right) signify good point candidates from the best (red) to second best (green) and third best (blue). Symbols i, j and i', j' denote the coordinate pairs before and after (') parameterization. From Publication IV.

At the first iteration when no edge points yet exist, the parameterization plane is set to middle of the volume. The tip the of chin, the parts of zygomatic bones are usually the first areas to get segmented and these areas extend further during subsequent iterations.

The parameterization defines the search space of voxels that are estimated with

$$E_{\text{ext (node)}} = a \left| \frac{\|\mathbf{g}_i\|}{\left(\frac{I_{\text{diff}}}{r}\right)} - 1 \right| + b \left| \frac{I(\mathbf{p}_i + j \cdot \mathbf{n}_i) - I(\mathbf{p}_i)}{I_{\text{diff}}} \right| + c \left| 1 - \cos^{-1} \left(\frac{\mathbf{g}_i \cdot \mathbf{n}_i}{\|\mathbf{g}_i\|} \right) \right| \quad (12)$$

for edge energy. There \mathbf{g}_i is the image gradient at voxel I , I_{diff} is the mean intensity difference between DC-clustering labeled voxels for the bone and soft tissue, r the radius of the Gaussian kernel and a , b and c weight parameters to be learned from the training set. The energy function favors points that have a strong gradient (term a), have an intensity drop similar to estimated drop from bone to soft tissue (term b) and have gradient orientation close to the normal orientation \mathbf{n}_i of the parameterized point \mathbf{p}_i (term c). The estimation of $E_{\text{ext(node)}}$ along a line profile is illustrated in Figure 20.

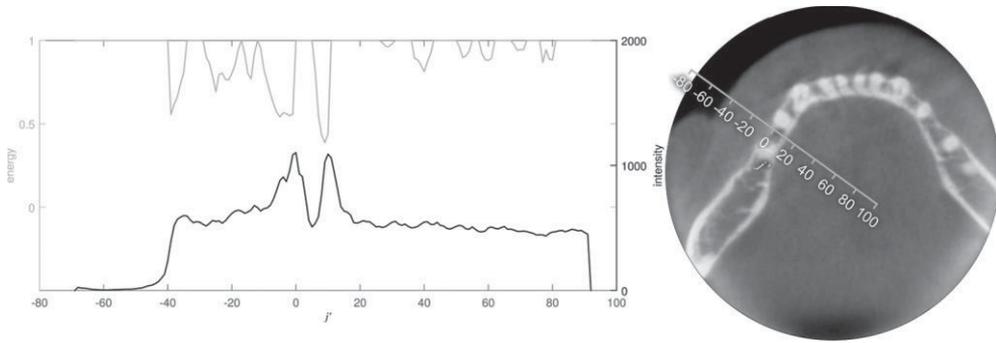


Figure 20. Energy minimization. Points of low energy (around $j'=-5$ and $j'=10$) correspond to high intensity gradients (left). These match the bone to soft tissue edges on image (right). From Publication IV.

Three potential edge points on the voxels of lowest energy per line segment are selected (Figure 19). In the following step the edge points of neighboring segments are connected with triangles and $E_{\text{ext}(\text{node})}$, estimated for all voxels that cross the triangles and the average is taken. Triangles that have the minimum energy are retained and those that exceed threshold energy are discarded. This step should be able to grow patches of connected surfaces where areas of consistent edges exist in the image (Figure 21).

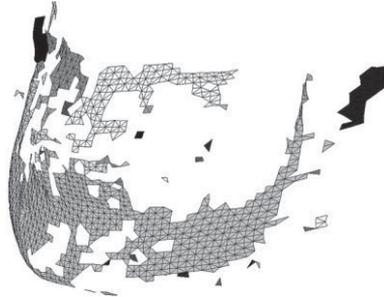


Figure 21. Surface patches covering the facial bones before merging. From Publication IV.

The result of the patch growing step is a surface that contains holes and discontinuities. These can be due to natural openings of the skull such as nose holes, missing teeth, mouth being open or image corruption from metal artefacts. These patches are connected by taking the largest patch in size and connecting all non-overlapping patches from largest to smallest. This step yields a surface that connects a maximum of one node per line segment. The resulting surface follows the minimum energy accurately but the holes remain. A standard thin-plate spline with smoothing

$$E_{\text{ext(spline)}}(f) = w \sum_{i=1}^N \|j - f(i)\|_2 + (1 - w)R(f) \quad (13)$$

is used to interpolate over the holes and to give the surface smoother appearance (Figure 22). A thin plate function is a weighted combination of the error sum term between the minimizer f and the smoothing term $R(f)$ typically composed of partial derivatives of f (Bookstein, 1989).

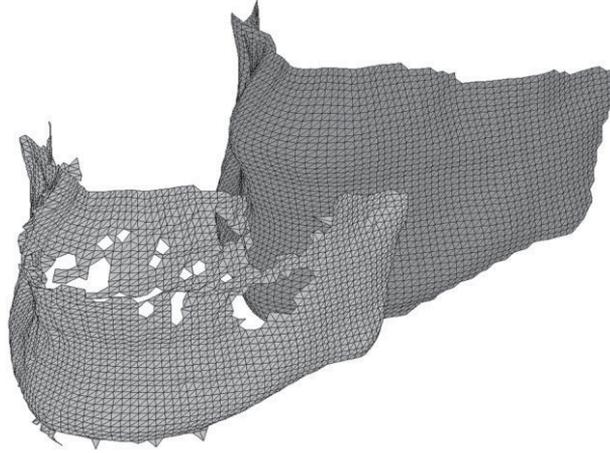


Figure 22. The facial bones after merging (foreground) and thin plate spline smoothing (background). From Publication IV.

This idea of using a spline is similar to what was presented by (Ahmad et al., 2012) although developed independently.

3.5.3 MR-HIFU pipeline

The motivation for building the MR-HIFU pipeline was born in a research setting independent from the previous works. However, the problems of segmenting facial bones in CBCTs and Gadolinium-contrasted PVs (the tumors) in MRIs have one common property: we want to find edges that drop from bright to dark. This enabled us to exploit much of the work put in for the NBVT and CBCT pipelines for the MR-HIFU.

The CBCT pipeline required a parameterization to fix the frame of reference where the segmentation was to be carried out. It used a simple, open surface (plane) bent along the shape of the face. This same approach would not work with PV for two reasons, it can take unpredictable shapes and it is, at least to some extent,

attached to the healthy organ. This means that the exterior shape of the PV is closer to a closed, sphere-like topology (genus-0) than an open surface. This lead us to change the strategy from the planar arc-length parameterization used for the facial bones to spherical parameterization presented in (Brechtbühler et al., 1995), (Székely et al., 1996) and (Shen & Davatzikos, 2000).

Apart from the parameterization strategy, the base elements from the CBCT pipeline for segmenting the PV were kept the same. The energy function was simplified to

$$\mathbf{g}_{i,j} = \nabla I(\mathbf{p}_i + j \cdot \mathbf{n}_i) \quad (14)$$

$$E_{\text{total}} = E_{\text{ext (hifu)}} = -\frac{1}{N} \sum_i \mathbf{g}_{i,0} \cdot \mathbf{n}_i \quad (15)$$

with the E_{int} component dropped altogether. $E_{\text{ext(hifu)}}$ is simply the sum of the dot products of the image gradient $\mathbf{g}_{i,0}$ and mesh normal vector \mathbf{n}_i at node i . With this change the PV segmentation relied more in the spline smoothing step to keep the parameterization consistent through iterations.

Solving the PV segmentation was only a part of the whole MR-HIFU problem. Segmenting the NPV was equally, if not more important. The PV segmentation provided a convenient mask inside which the binary PV vs. NPV segmentation could be performed. After benchmarking a number of candidates we ended up using EM-segmentation (Moon, 1996), (Van Leemput et al., 1999) for solving the bulk binary classification and Markov Random-Fields (MRFs) (Van Leemput et al., 1999), (Salli et al., 2001) to refine the segmentation result locally.

The EM/MRF segmentation yielded a binary classification that sometimes revealed imperfections in the original PV segmentation. These appeared in areas where the dark tissue type penetrated the PV, such as blood veins, the necrosis puncturing the tumor edge or imperfect (secondary) Gadolinium perfusion due to image timing inaccuracies. These imperfections would show as NPV leaks outside binary PV containment. As a final step of the segmentation pipeline these leaks were capped by running the same PV segmentation once more with the original PV surface as starting point on the binary PV vs. NPV mask.

4 RESULTS

Our algorithms were optimized to minimize the mean squared distance from the produced segmentation against a given reference. We were able to reach 0.32 ± 0.65 (avg \pm std) mm mean MSD for the NBVT (Publication II) and 0.50 ± 1.01 mm for the CBCT (Publication IV) pipelines against manually drawn surfaces. The CBCT pipeline was also able to reach 92% mean coverage of the nodes in reference surface meshes with under a minute mean execution time per volume.

Two summary tables of results were collected from the studies reviewed in section 2.4.1. The studies in Table 2 use either cone- or narrow beam imaging, target mandible, maxilla or both and report an MSD metric for their results. These criteria make the results in the table a natural comparison against Publications I- IV. Studies that present relevant methodology but use either CT or report only Dice similarity coefficient (DSC) are in Table 3. Thus the results summarized in this table are harder to compare against Publications I- IV but since they present some of the latest advances in the field, they are included for reference. The results in Table 2 that have reported MSD are also presented visually in a timeline in Figure 23. It should be noted that the precise definition on the MSD is not uniform between the reference articles but include an interpretation of average error between the target and reference in millimeters.

The reviewed articles typically estimate the MSD per sample and give the mean and standard deviation over the entire dataset and sometimes for a subset held out for the validation. In Table 2, the MSDs for the whole data was because of its better availability. The median is used in comparisons over publications. It was chosen over mean to mitigate the effect of outliers in the comparisons.

Our results compare well against the reviewed literature. The achieved MSDs of 0.32 mm for the mandible and 0.50 mm for the facial bones fall in line of what has been reported. The result of 0.32 mm (NBVT) is below the median of 0.5 mm (results in Table 2 for targets that do not include the maxilla). The result of 0.5 mm (CBCT) is above for the median of 0.25 mm (results in Table 2 that include both mandible and maxilla). The median MSD for all studies Table 2 excluding our work is 0.42 mm.

Table 2. Dental segmentation methods for CBCT that report an MSD value or similar error metric summarized

source	MSD (mm)	n	modality	principle	ROI	notes
Lamecker et al., 2006	~0.19	15	CBCT	SSM	mandible, no teeth	first segmentation for CBCT
<i>Publication I</i>	0.36 ± 0.01	3	NBVT	ACM	mandible segment	
<i>Publication II</i>	0.32 ± 0.65	9+4	NBVT	ACM	mandible segment	
<i>Publication III</i>	0.57 ± 0.16	14	CBCT	ACM	mandible, no teeth	the surface model includes tooth orientation
Kainmueller et al., 2009	0.5 ± 0.1	106	CBCT	SSM	mandible, no teeth	mandibular nerve included, single clinical site
Gollmer & Buzug, 2012	0.5 ± 0.0	30	CBCT	SSM	mandible, no teeth	unique shape priors included
Chang et al., 2013	0.25 ± 0.2	19	CBCT	WDM	maxilla (anterior)	only part of anterior wall of maxilla segmented
Wang et al., 2014	0.65 ± 0.19	15	CBCT	MA	mandible, maxilla	error given only for mandible, 5 h running time
<i>Publication IV</i>	0.50 ± 1.01	19+20	CBCT	patches	mandible & maxilla	
Wang et al., 2016	0.42 ± 0.15	30	CBCT	MA & RF	mandible & maxilla	error only for mandible, 20 min running time
Abdolali et al., 2017	0.71 ± 0.09	120	CBCT	SSM	mandible & canal	method developed primarily for the canal
Minnema et al., 2019	0.40 ± 0.12	20	CBCT	CNN	mandible & maxilla	comparison of ACM and U-Net
Vaitiekūnas et al., 2019	0.057	20+20	CBCT	Otsu	mandible & maxilla	20 patients, 20 pre and 20 post-op. images

Table 3. Dental segmentation methods with methodological value, but developed primary for CT or the authors do not report results in MSD.

source	MSD (mm)	n	modality	principle	ROI	notes
Rueda et al., 2006	1.63	215	CT	AAM	mandible & others	first segmentation of the mandibular canal
Le et al., 2009	0.28 & 0.30	2	CT	graph cuts	mandible & maxilla	discusses maxilla to mandible separation
Lloréns et al., 2012	0.16 \pm 0.03	20	CT	FC	mandible & cortical	a follow-up work for Rueda et. al. 2006
Duy et al., 2012	< 1	43	CT	SSM	maxilla & teeth	a method for tooth separation presented
Zhang et al., 2017		77+30	CBCT & CT	CNN	mandible & maxilla	0.93 \pm 0.01 DSC reported
Fan et al., 2018	< 1	21	CBCT	watersheds	mandible	0.97 \pm 0.01 DSC reported
Torosdagli et al., 2019		50+48	CBCT & CT	CNN	mandible	0.94 DSC reported, dataset with deformities

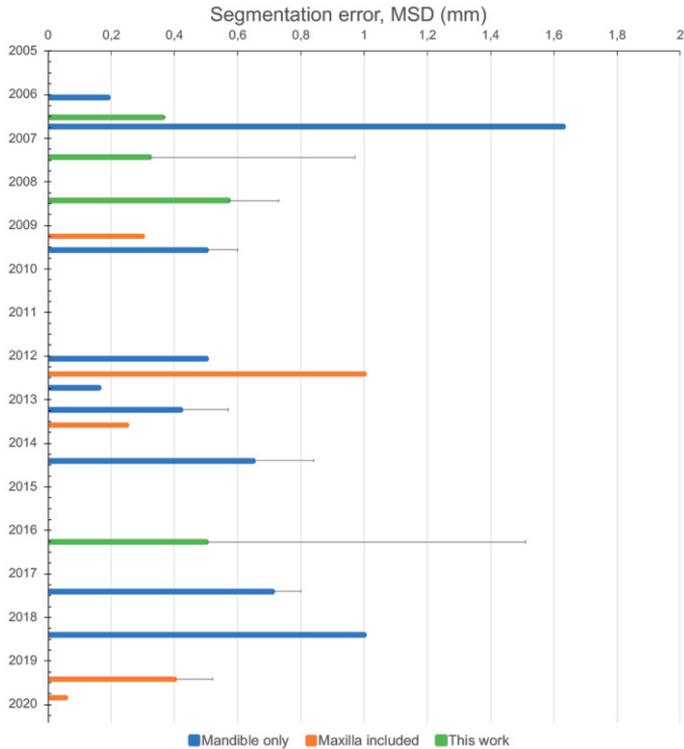


Figure 23. Progression of the dental segmentation error over time in reviewed literature. All authors have not provided the standard deviation of their reported error. The standard deviation is drawn as an upper whisker to the bar when available.

By only looking at the median error, the methods that segment only mandible (median of 0.5 mm) appears to be less accurate than methods that target both mandible and maxilla (median of 0.25 mm). This is counterintuitive since maxilla should be a more difficult target. A possible explanation for this is that the methods that target both are newer and benefit from later methodology and improved image quality. However, there are only four studies (Chang et al., 2013, Publication IV, Minnema et al., 2019 and Vaitiekūnas et al., 2019) that target both mandible and maxilla. The work by (Vaitiekūnas et al., 2019) appears to be an outlier with a MSD of only 0.057 mm.

Of the studies in Table 3. one could note the works by (Zhang et al., 2017) and (Torosdagli et al., 2019) for their larger than average sample sets ($n = 107$ and $n = 98$) and accurate results (DSCs of 0.93 and 0.94).

Objective comparisons between mandibular segmentation algorithms are complicated by fact that all of them were tested and validated on proprietary datasets.

Very few of the reviewed studies characterize the factors affecting the dataset image quality, mostly metal artefacts and the numbers and of types and pathologies in the samples. Exceptions of this are (Publication IV), (Minnema et al., 2019) and (Torosdagli et al., 2019). Table 2 also shows that studies with CBCT were done with relatively small datasets. The median sample size over all publications is a rather modest $n = 23$ with only (Kainmueller et al., 2009) and (Abdolali et al., 2017) exceeding the $n > 100$ mark.

Our MR-HIFU pipeline achieved the DSC error of 0.88. The review of the literature revealed only nine other original papers discussing the uterine fibroid segmentation. The DSC errors segmentation in these papers range from 0.80 in (Fallahi et al., 2011) to 0.88 (Militello et al., 2015). Our result of 0.88 is not directly comparable with the other work since it was estimated for the post-operative, non-perfused volume while the other studies typically report the result for the whole uterus including the tumor if present. Based on the literature review, Publication V is the only a method with results for the non-perfused volume segmentation. The mean execution time for the MR-HIFU pipeline was 1.6 minutes per volume.

It can be noted that the most accurate result (Militello et al., 2015) of DSC 0.88 in the review was validated with a rather small, $n=15$ dataset. The study by (Kurata et al., 2019) reported DSCs of 0.84 (subjects with a disease) and 0.78 (healthy subjects) with an $n = 122$ sized dataset.

5 DISCUSSION

In this work we developed segmentation methods for facial bones and uterine fibroids. The planned use for these was in clinical workflows of planning implants and updating MR-HIFU treatment plans during the therapy. Any segmentation method considered to be deployed in such a demanding environment needs to be robust and computationally efficient. Robustness is necessary to ensure the overall reliability of the results acquired with the method. A robust method is also less prone to breaking on images with artefacts or other factors degrading the quality. Better robustness adds value to the implant workflow by reducing the need for performing the planning steps on poor quality images manually. Computational efficiency speeds up the implant workflow that is still largely interactive. When less time is spent waiting for segmentations to complete, more operator time can be invested elsewhere. Real-time performance is much desired in MR-HIFU since the result is needed in guiding the therapy while the patient is sedated and positioned on the device.

The requirement for robustness was addressed early in the design by developing methods that are able to tolerate noise and piece together surfaces in images suffering from metal and inhomogeneity artefacts. The robustness was tested by validating the results on independent test sets (Publication IV and Publication V) More, the segmentation targets may only consist only partial and thus discontinuing edges. These are characteristic properties of both cone-beam computed tomography used in imaging facial bones and magnetic resonance imaging used for fibroids. The performance requirement was managed by selecting and developing segmentation pipeline components capable of completing rather in seconds than minutes. Our CBCT pipeline reached an s under minute and the MR-HIFU slightly over minute average execution times from start to finish even in their non-optimized implementation and using ordinary desktop computer hardware. Even though an early decision was made to concentrate only to fast algorithms, the CBCT pipeline was to maintain a similar accuracy of 0.50 mm of mean squared distance error against 0.42 mm to methods in the reviewed literature including the latest in the field. Thorough comparisons of the accuracy could not be made for MR-HIFU since it

remains the only known method on segmenting necrosed, non-perfused volumes of uterine fibroids.

The development and validation of image segmentation methods are a function of and constrained by the availability and properties of the input data. This is especially true with modalities that use x-rays, where bone segmentation from clean, artefact-free images can be almost trivial and again extremely difficult in the presence of artefacts. This might partly explain why the complexity of the proposed solutions for dental image segmentation range from selecting a single, global threshold (Miet Loubele et al., 2006) to training a deep learning neural network of 37 million parameters (Minnema et al., 2019). Thresholding has again been suggested as late as 2019 (Vaitiekūnas et al., 2019). Only very few of the articles reviewed for this work actually describe the state, type and prevalence of the quality issues in the input data. Based on the wide variety of the solutions proposed, it could be speculated that there is significant variation in both data contents and quality. The main trends in medical image segmentation from active contour models through multi-atlas methods to deep learning are apparent in the CBCT domain as well. Also, CBCT is still an evolving modality. Advances in improving the image quality have been achieved to this day and are likely still to be made.

The success or failure of our and similar segmentation methods depend heavily on how the energy function that is used to drive the algorithm towards better accuracy is defined. The form of the functions developed for the subproblems in this work are a result of going through iterations of collecting ideas, implementations, experimentation and going back to collect better ideas. At the same time more data was collected and even modalities were changed along the way. Even of the development of the energy functions is done thoroughly, it is mostly a product of the designer and less is learned from the data directly. In this work the development process was successful in producing simple, generalizable and explainable methods. When the input sample sites like ours are rather small it is also quite natural to give more influence on the designer. However, when the image sample sets grow to hundreds let alone thousands, the performance of methods designed this way start to eventually stagnate.

A good part of the emergence of deep learning (Lecun et al., 2015) as the dominant family of methods for image analysis and machine learning in general can be attributed to the availability of massive datasets such as ImageNet (Russakovsky et al., 2015) and the associated competition-form challenges. One very desirable property of deep learning for medical image analysis (Shen et al., 2017) is that it takes the burden of writing the energy function and engineering its input

features away from the designer. Methods of deep learning are in a good position to utilize the massive accumulation of image data by training larger models that should be more capable capturing the variation in the data. Despite these benefits the neural networks still require the designer to make decisions on items such as pre-processing, network topology and the loss function. These necessary components of the image processing pipeline are hard to integrate as trainable hyperparameters. Deep learning networks are still computationally expensive to train and hardware that can handle large models of high-resolution volumetric images can carry a significant investment cost. Deep networks can also get distracted by changes in the image collection protocol or other due to other variations (Der Sarkissian et al., 2019). A classic example of this to train on a large, public domain data set and test on smaller set drawn from another distribution (Mårtensson et al., 2020) such as hospital's own, proprietary data. There are means to combat this (Fang et al., 2020), but they require that problem is first recognized and then addressed. There are convincing examples of transfer learning, from the capability of an algorithm to be trained in a non-medical domain to be successfully used on medical images (Raghu et al., 2019). Scanning protocols are still very much derived from the needs of (individual) consulting radiologists and not standardized for a machine learning algorithm. This will probably remain so in the near future. Therefore, the algorithms will be also required to adapt to different domains.

Despite the dominance of deep learning in a wide spectrum of image classification and segmentation problems they have been proposed for dental CBCT only quite recently (Zhang et al., 2017), (Torosdagli et al., 2019), (Minnema et al., 2019). The reason for this late onset might be in part due to the sparsity of annotated, generally good quality and public datasets with well specified protocols in decent sizes. Collecting sufficiently large CBCT data sets is hard because it is not possible to expose healthy volunteers to ionizing radiation and a full anonymization of patients from tomographic scans of the facial structures is not generally possible. There are remarkably well working methods to synthesize data including generative adversarial networks (GANs) (Karras et al., 2018) but no coordinated effort to use them to generate simulated training sets for dental volumetric imaging appears to have been made.

The literature review revealed many indications of the benefits of deep learning in CBCT segmentation. The validation work (Minnema et al., 2019) suggest that convolutional neural networks (CNNs) perform better than active contour models on metal artefact corrupted data. The margin is small but significant. The adoption of CNNs by (Zhang et al., 2017) over their multi-atlas method (Wang et al., 2016)

brought accuracy gains as well. Finally, the thorough validation of (Torosdagli et al., 2019) on data specifically selected to contain samples of subjects with facial deformities with good results is strong argument for the flexibility of deep learning. These results combined suggest that CNNs are currently the best performing strategy for mandibular and maxillary CBCT segmentation.

By their definition, CNNs use convolutional kernels for extracting the primitive characteristics of shapes for the purpose of classifying objects in images. An example this is the very popular U-net (Ronneberger et al., 2015). All of the end results presented in this work can be derived from convolutions but this is possibly only indirectly. The kernels work best for a data that is presented in a fixed geometry such as an image voxel coordinate grid and worse, or not at all, in a deforming topology such as surface mesh. Thus convolutions are not inherently good in fitting surfaces.

To solve the surface fitting problem in its native form with deep learning would require departure from the Euclidean voxel grid to a space that is somehow parameterized. Segmentations in parameterized spaces were used extensively also in this work. A move away from the Euclidean space probably require abandoning the traditional formulation of the convolutions as well. A fairly recent introduction of geometric deep learning by (Bronstein et al., 2017) or more generally the graph neural networks (Wu et al., 2020) could provide some of the necessary tools for this. Some of the published works provide accurate characterization of shapes for the classification (Fey et al., 2018) and segmentation to sub-meshes (Boscaini & Poiesi, 2019). These ideas were used in the research by (Zhang et al., 2017) and especially (Torosdagli et al., 2019) although they handle the segmentation and the surface characterization problems separately. Graph networks could also solve a potential weakness in how the CBCT and MR-HIFU pipelines would maintain the integrity of the parameterized space throughout iterations.

What appears to be still largely unstudied, is how to best embed or encode the properties of the image intensities on surrounding the surface mesh with a deep network. One potential approach would be to learn the facial shape correspondences with anisotropic convolutional networks (Boscaini et al., 2016) and embed the grayscale intensities in regular convolutions along corresponding shapes. To solve this problem efficiently and with a feasibly sized network with relatively sparse availability of annotated input data is no easy task. This approach would also require solving the same point correspondence problems that largely forced us to abandon the multi-atlas segmentation approach attempted prior to the patch-based CBCT pipeline development.

6 CONCLUSION

This work presented segmentation methods for facial bones and muscle tumors. These methods were designed to produce accurate surface reconstructions and since intended for clinical workflow they were built to be both robust and fast. We were able to develop methods that matched the accuracy of published methods for similar targets. This work demonstrates the usefulness of simple and generic ideas with explainable methodology that can cross application domains.

7 REFERENCES

- Abdolali, F., Zoroofi, R. A., Abdolali, M., Yokota, F., Otake, Y., & Sato, Y. (2017). Automatic segmentation of mandibular canal in cone beam CT images using conditional statistical shape model and fast marching. *International Journal of Computer Assisted Radiology and Surgery*, *12*(4), 581–593. <https://doi.org/10.1007/s11548-016-1484-2>
- Ahmad, R. R., Ghazali, N., Rambely, A. S., Din, U. K. S., & Hassan, N. (2012). Application of cubic spline in the implementation of braces for the case of a child. *Journal of Mathematics and Statistics*, *8*(1), 144–149. <https://doi.org/10.3844/jmssp.2012.144.149>
- Aiello, M., Cavaliere, C., D’Albore, A., & Salvatore, M. (2019). The Challenges of Diagnostic Imaging in the Era of Big Data. *Journal of Clinical Medicine*, *8*(3), 316. <https://doi.org/10.3390/jcm8030316>
- Akhoondali, H., Zoroofi, R. A., & Shirani, G. (2009). Fully automatic extraction of panoramic dental images from CT-scan volumetric data of the head. *Journal of Applied Sciences*, *9*(11), 2106–2114. <https://doi.org/10.3923/jas.2009.2106.2114>
- AL-Allaf, O. N. A. (2014). Review of Face Detection Systems Based Artificial Neural Networks Algorithms. *The International Journal of Multimedia & Its Applications*, *6*(1), 1–16. <https://doi.org/10.5121/ijma.2013.6101>
- Alexander, A., McGill, M., Tarasova, A., Ferreira, C., & Zurkiya, D. (2019). Scanning the Future of Medical Imaging. *Journal of the American College of Radiology*, *16*(4), 501–507. <https://doi.org/10.1016/j.jacr.2018.09.050>
- Aljabar, P., Heckemann, R. A., Hammers, A., Hajnal, J. V., & Rueckert, D. (2009). Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage*, *46*(3), 726–738. <https://doi.org/10.1016/j.neuroimage.2009.02.018>
- Andersen, A. H. (1989). Algebraic Reconstruction in CT from Limited Views. In *IEEE Transactions on Medical Imaging* (Vol. 8, Issue 1, pp. 50–55). <https://doi.org/10.1109/42.20361>
- Arifin, A. Z., & Asano, A. (2006). Image segmentation by histogram thresholding using hierarchical cluster analysis. *Pattern Recognition Letters*, *27*(13), 1515–1521. <https://doi.org/10.1016/j.patrec.2006.02.022>
- Arun, K. S., Huang, T. S., & Blostein, S. D. (1987). Least-Squares Fitting of Two 3-D Point Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-9*(5), 698–700. <https://doi.org/10.1109/TPAMI.1987.4767965>

- BachCuadra, M., Duay, V., & Thiran, J. P. H. (2015). Atlas-based segmentation. In *Handbook of Biomedical Imaging: Methodologies and Clinical Research* (pp. 221–244). Springer US. https://doi.org/10.1007/978-0-387-09749-7_12
- Bae, M., Park, J. W., & Kim, N. (2019). Semi-automatic and robust determination of dental arch form in dental cone-beam CT with B-spline approximation. *Computer Methods and Programs in Biomedicine*, 172, 95–101. <https://doi.org/10.1016/j.cmpb.2019.02.013>
- Baillard, C., Hellier, P., & Barillot, C. (2001). Segmentation of brain 3D MR images using level sets and dense registration. *Medical Image Analysis*, 5(4), 185–194. [https://doi.org/10.1016/S1361-8415\(01\)00039-1](https://doi.org/10.1016/S1361-8415(01)00039-1)
- Barandiaran, I., MacÍa, I., Berckmann, E., Wald, D., Dupillier, M. P., Paloc, C., & Graña, M. (2009). An automatic segmentation and reconstruction of mandibular structures from CT-data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5788 LNCS, 649–655. https://doi.org/10.1007/978-3-642-04394-9_79
- Barber, C. B., Dobkin, D. P., & Huhdanpaa, H. (1996). The Quickhull Algorithm for Convex Hulls. *ACM Transactions on Mathematical Software*, 22(4), 469–483. <https://doi.org/10.1145/235815.235821>
- Beyerer, J., León, F. P., & Frese, C. (2015). Machine vision: Automated visual inspection: Theory, practice and applications. In *Machine Vision: Automated Visual Inspection: Theory, Practice and Applications*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-47794-6>
- Bloch, P., & Udupa, J. K. (1983). Application of Computerized Tomography to Radiation Therapy and Surgical Planning. *Proceedings of the IEEE*, 71(3), 351–355. <https://doi.org/10.1109/PROC.1983.12593>
- Bookstein, F. L. (1989). Principal warps: thin-plate splines and the decomposition of deformations. *Pattern Analysis and Machine Intelligence, IEEE Transactions On*, 11(6), 567–585. <https://doi.org/10.1109/34.24792>
- Boscaini, D., Masci, J., Bronstein, M., & Lugano, U. (2016). *Learning shape correspondence with anisotropic convolutional neural networks*.
- Boscaini, D., & Poiesi, F. (2019). 3D shape segmentation with geometric deep learning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11751 LNCS, 454–465. https://doi.org/10.1007/978-3-030-30642-7_41
- Boykov, Y., & Funka-Lea, G. (2006). Graph cuts and efficient N-D image segmentation. *International Journal of Computer Vision*, 70(2), 109–131. <https://doi.org/10.1007/s11263-006-7934-5>
- Brechbühler, C., Gerig, G., & Kübler, O. (1995). Parametrization of Closed Surfaces for 3-D Shape Description. *Comput. Vis. and Imag. Underst.*, 61(2), 154–170.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

- <https://doi.org/10.1023/A:1010933404324>
- Bronstein, M. M., Bruna, J., Lecun, Y., Szlam, A., & Vandergheynst, P. (2017). Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4), 18–24. <https://doi.org/10.1109/MSP.2017.2693418>
- Buades, A., Coll, B., & Morel, J. M. (2005). A non-local algorithm for image denoising. *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, II*, 60–65. <https://doi.org/10.1109/CVPR.2005.38>
- Burn, P. R., McCall, J. M., Chinn, R. J., Vashisht, A., Smith, J. R., & Healy, J. C. (2000). Uterine fibroleiomyoma: MR imaging appearances before and after embolization of uterine arteries. *Radiology*, 214(3), 729–734. <https://doi.org/10.1148/radiology.214.3.r00fe07729>
- Cederlund, A., Kalke, M., & Welander, U. (2014). Volumetric tomography – a new tomographic technique for panoramic units. <Http://Dx.Doi.Org/10.1259/Dmfr/16577933>. <https://doi.org/10.1259/DMFR/16577933>
- Chang, Y. B., Xia, J. J., Yuan, P., Kuo, T. H., Xiong, Z., Gateno, J., & Zhou, X. (2013). 3D segmentation of maxilla in cone-beam computed tomography imaging using base invariant wavelet active shape model on customized two-manifold topology. *Journal of X-Ray Science and Technology*, 21(2), 251–282. <https://doi.org/10.3233/XST-130369>
- Chanwimaluang, T., Sotthivirat, S., & Sinthupinyo, W. (2008). Automated dental arch detection using computed tomography images. *International Conference on Signal Processing Proceedings, ICSP*, 737–740. <https://doi.org/10.1109/ICOSP.2008.4697235>
- Chindasombatjaroen, J., Kakimoto, N., Murakami, S., Maeda, Y., & Furukawa, S. (2011). Quantitative analysis of metallic artifacts caused by dental metals: Comparison of cone-beam and multi-detector row CT scanners. *Oral Radiology*, 27(2), 114–120. <https://doi.org/10.1007/s11282-011-0071-z>
- Chuang, Y. J., Doherty, B. M., Adluru, N., Chung, M. K., & Vorperian, H. K. (2018). A Novel Registration-Based Semiautomatic Mandible Segmentation Pipeline Using Computed Tomography Images to Study Mandibular Development. *Journal of Computer Assisted Tomography*, 42(2), 306–316. <https://doi.org/10.1097/RCT.0000000000000669>
- Çiftçi, M. E., Aktan, A. M., İşman, Ö., & Yıldırım, E. (2016). Relationship between CBCT and panoramic images of the morphology and angulation of the posterior mandibular jaw bone. *Surgical and Radiologic Anatomy*, 38(3), 313–320. <https://doi.org/10.1007/s00276-015-1553-1>
- Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *Pattern Analysis and Machine Intelligence, IEEE Transactions On*, 23(6), 681–685. <https://doi.org/10.1109/34.927467>

- Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active Shape Models-Their Training and Application. *Computer Vision and Image Understanding*, 61(1), 38–59. <https://doi.org/http://dx.doi.org/10.1006/cviu.1995.1004>
- Coupé, P., Manjón, J. V., Fonov, V., Pruessner, J., Robles, M., & Collins, D. L. (2011). Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage*, 54(2), 940–954. <https://doi.org/10.1016/j.neuroimage.2010.09.018>
- Cuadros Linares, O., Bianchi, J., Raveli, D., Batista Neto, J., & Hamann, B. (2019). Mandible and skull segmentation in cone beam computed tomography using super-voxels and graph clustering. *Visual Computer*, 35(10), 1461–1474. <https://doi.org/10.1007/s00371-018-1511-0>
- Cucchiara, R., Lamma, E., & Sansoni, T. (2004). An image analysis approach for automatically re-orienting CT images for dental implants. *Computerized Medical Imaging and Graphics*, 28(4), 185–201. <https://doi.org/10.1016/j.compmedimag.2003.12.004>
- Curless, B., & Levoy, M. (1996). A volumetric method for building complex models from range images. *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1996*, 303–312. <https://doi.org/10.1145/237170.237269>
- Der Sarkissian, H., Lucka, F., van Eijnatten, M., Colacicco, G., Coban, S. B., & Batenburg, K. J. (2019). A cone-beam X-ray computed tomography data collection designed for machine learning. *Scientific Data*, 6(1), 215. <https://doi.org/10.1038/s41597-019-0235-y>
- Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3), 297–302. <https://doi.org/10.2307/1932409>
- Dinov, I. D. (2016). Volume and value of big healthcare data. *Journal of Medical Statistics and Informatics*, 4(1), 3. <https://doi.org/10.7243/2053-7662-4-3>
- Duy, N. T., Lamecker, H., Kainmueller, D., & Zachow, S. (2012). Automatic detection and classification of teeth in CT data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7510 LNCS, 609–616. https://doi.org/10.1007/978-3-642-33415-3_75
- Elhelf, I. A. S., Albahar, H., Shah, U., Oto, A., Cressman, E., & Almekawy, M. (2018). High intensity focused ultrasound: The fundamentals, clinical applications and research trends. In *Diagnostic and Interventional Imaging* (Vol. 99, Issue 6, pp. 349–359). Elsevier Masson SAS. <https://doi.org/10.1016/j.diii.2018.03.001>
- Fallahi, A., Pooyan, M., Ghanaati, H., Oghabian, M. A., Khotanlou, H., Shakiba, M., Jalali, A. H., & Firouznia, K. (2011). Uterine segmentation and volume measurement in uterine fibroid patients' MRI using Fuzzy C-Mean algorithm and morphological operations. *Iranian Journal of Radiology*, 8(3),

- 150–156. <https://doi.org/10.5812/kmp.iranradiol.17351065.3142>
- Fallahi, A., Pooyan, M., Khotanlou, H., Hashemi, H., Firouznia, K., & Oghabian, M. A. (2010). Uterine fibroid segmentation on multiplan MRI using FCM, MPFCM and morphological operations. *ICCET 2010 - 2010 International Conference on Computer Engineering and Technology, Proceedings*, 7. <https://doi.org/10.1109/ICCET.2010.5485920>
- Fan, Y., Beare, R., Matthews, H., Schneider, P., Kilpatrick, N., Clement, J., Claes, P., Penington, A., & Adamson, C. (2018). Marker-based watershed transform method for fully automatic mandibular segmentation from low-dose CBCT images. *BioRxiv*, 397166. <https://doi.org/10.1101/397166>
- Fang, T., Riken, K. & Riken, G. N., Lu, N., & Sugiyama, M. (2020). *Rethinking Importance Weighting for Deep Learning under Distribution Shift*. <https://arxiv.org/pdf/2006.04662.pdf>
- Fey, M., Lensen, J. E., Weichert, F., & Müller, H. (2018). *SplineCNN: Fast Geometric Deep Learning with Continuous B-Spline Kernels*. https://github.com/rusty1s/pytorch_geometric
- Gollmer, S. T., & Buzug, T. M. (2012). Fully automatic shape constrained mandible segmentation from cone-beam CT data. *Proceedings - International Symposium on Biomedical Imaging*, 1272–1275. <https://doi.org/10.1109/ISBI.2012.6235794>
- Haddad, R. A., & Akansu, A. N. (1991). A Class of Fast Gaussian Binomial Filters for Speech and Image Processing. *IEEE Transactions on Signal Processing*, 39(3), 723–727. <https://doi.org/10.1109/78.80892>
- Han, B., Chen, L., Cai, Z., Pu, F. L. D., Li, S., & Fan, Y. (2011). An automatic method of synthesizing panoramic radiograph by unwrapping dental CT image. *Proceedings 2011 International Conference on Mechatronic Science, Electric Engineering and Computer, MEC 2011*, 1094–1096. <https://doi.org/10.1109/MEC.2011.6025657>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Indraswari, R., Zainal Arifin, A., Suciati, N., Renwi Astuti, E., & Kurita, T. (2019). Automatic Segmentation of Mandibular Cortical Bone on Cone-Beam CT Images Based on Histogram Thresholding and Polynomial Fitting. *International Journal of Intelligent Engineering and Systems*, 12(4). <https://doi.org/10.22266/ijies2019.0831.13>
- Jegou, S., Drozdal, M., Vazquez, D., Romero, A., & Bengio, Y. (2017). The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2017-July, 1175–1183.

- <https://doi.org/10.1109/CVPRW.2017.156>
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). Review FSL. *NeuroImage*, 62(2), 782–790. <https://doi.org/10.1016/j.neuroimage.2011.09.015>
- Jones, D. R., Perttunen, C. D., & Stuckman, B. E. (1993). Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Applications*, 79(1), 157–181. <https://doi.org/10.1007/BF00941892>
- Kainmueller, D., Lamecker, H., Seim, H., Zinser, M., & Zachow, S. (2009). Automatic extraction of mandibular nerve and bone from cone-beam CT data. *Medical Image Computing and Computer-Assisted Intervention: MICCAI ...International Conference on Medical Image Computing and Computer-Assisted Intervention*, 12(Pt 2), 76–83. https://doi.org/10.1007/978-3-642-04271-3_10
- Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). Progressive growing of GANs for improved quality, stability, and variation. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*.
- Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, 1, 321–331. <https://doi.org/10.1007/BF00133570>
- Kennedy, J. E. (2005). High-intensity focused ultrasound in the treatment of solid tumours. In *Nature Reviews Cancer* (Vol. 5, Issue 4, pp. 321–327). Nature Publishing Group. <https://doi.org/10.1038/nrc1591>
- Khotanlou, H., Fallahi, A., Oghabian, M. A., & Pooyan, M. (2014). Segmentation of uterine fibroid on MR images based on chan-veve level set method and shape prior model. *Biomedical Engineering - Applications, Basis and Communications*, 26(2). <https://doi.org/10.4015/S1016237214500306>
- Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671–680. <https://doi.org/10.1126/science.220.4598.671>
- Köhler, M. O., Mougenot, C., Quesson, B., Enhholm, J., Le Bail, B., Laurent, C., Moonen, C. T. W., & Ehnholm, G. J. (2009). Volumetric HIFU ablation under 3D guidance of rapid MRI thermometry. *Medical Physics*, 36(8), 3521–3535. <https://doi.org/10.1118/1.3152112>
- Kotcheff, A. C. W., & Taylor, C. J. (1998). Automatic construction of eigenshape models by direct optimization. *Medical Image Analysis*, 2(4), 303–314. [https://doi.org/10.1016/s1361-8415\(98\)80012-1](https://doi.org/10.1016/s1361-8415(98)80012-1)
- Kurata, Y., Nishio, M., Kido, A., Fujimoto, K., Yakami, M., Isoda, H., & Togashi, K. (2019). Automatic segmentation of the uterus on MRI using a convolutional neural network. *Computers in Biology and Medicine*, 114, 103438. <https://doi.org/10.1016/j.combiomed.2019.103438>
- Lamecker, H., Zachow, S., Wittmers, A., Weber, B., Hege, H.-C., Elsholtz, B., &

- Stiller, M. (2006). Automatic Segmentation of Mandibles in Low-Dose CT-Data. *Intern. J. of Comput. Ass. Radio.and Surg.*, 1(1), 393–395.
- Le, B. H., Deng, Z., Xia, J., Chang, Y. B., & Zhou, X. (2009). An interactive geometric technique for upper and lower teeth segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5762 LNCS(PART 2), 968–975. https://doi.org/10.1007/978-3-642-04271-3_117
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. In *Nature* (Vol. 521, Issue 7553, pp. 436–444). Nature Publishing Group. <https://doi.org/10.1038/nature14539>
- Lénárd, Z. M., McDannold, N. J., Fennessy, F. M., Stewart, E. A., Jolesz, F. A., Hynynen, K., & Tempny, C. M. C. (2008). Uterine leiomyomas: MR imaging-guided focused ultrasound surgery-imaging predictors of success 1. *Radiology*, 249(1), 187–194. <https://doi.org/10.1148/radiol.2491071600>
- Liang, X., Lambrichts, I., Sun, Y., Denis, K., Hassan, B., Li, L., Pauwels, R., & Jacobs, R. (2010). A comparative evaluation of Cone beam computed tomography (CBCT) and multi-slice CT (MSCT). Part II: On 3D model accuracy. *European Journal of Radiology*, 75(2), 270–274. <https://doi.org/10.1016/j.ejrad.2009.04.016>
- Littleton, J. T., & Durizch Littleton, M. L. (1996). Conventional Tomography. In R. Gagliardi & B. L. McClennan (Eds.), *A history of the Radiological Sciences* (pp. 369–401). American Roentgen Ray Society. <https://www.arrs.org/ARRSLIVE/HRS>
- Lloréns, R., Naranjo, V., López, F., & Alcañiz, M. (2012). Jaw tissues segmentation in dental 3D CT images using fuzzy-connectedness and morphological processing. *Computer Methods and Programs in Biomedicine*, 108(2), 832–843. <https://doi.org/10.1016/j.cmpb.2012.05.014>
- Lorenson, W. E., & Cline, H. E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1987*, 163–169. <https://doi.org/10.1145/37401.37422>
- Lötjönen, J., & Mäkelä, T. (2001). Elastic matching using a deformation sphere. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2208, 541–548. https://doi.org/10.1007/3-540-45468-3_65
- Lötjönen, J., Reissman, P. J., Magnin, I. E., & Katila, T. (1999). Model extraction from magnetic resonance volume data using the deformable pyramid. *Medical Image Analysis*, 3(4), 387–406. [https://doi.org/10.1016/S1361-8415\(99\)80031-0](https://doi.org/10.1016/S1361-8415(99)80031-0)
- Lötjönen, J., Wolz, R., Koikkalainen, J., Thurfjell, L., Waldemar, G., Soininen, H., & Rueckert, D. (2010). Fast and robust multi-atlas segmentation of brain magnetic resonance images. *NeuroImage*, 49(3), 2352–2365.

- <https://doi.org/10.1016/j.neuroimage.2009.10.026>
- Loubele, M., Bogaerts, R., Van Dijck, E., Pauwels, R., Vanheusden, S., Suetens, P., Marchal, G., Sanderink, G., & Jacobs, R. (2009). Comparison between effective radiation dose of CBCT and MSCT scanners for dentomaxillofacial applications. *European Journal of Radiology*, 71(3), 461–468. <https://doi.org/10.1016/j.ejrad.2008.06.002>
- Loubele, M., Jacobs, R., Maes, F., Denis, K., White, S., Coudyzer, W., Lambrechts, I., Van Steenberghe, D., & Suetens, P. (2008). Image quality vs radiation dose of four cone beam computed tomography scanners. *Dentomaxillofacial Radiology*, 37(6), 309–318. <https://doi.org/10.1259/dmfr/16770531>
- Loubele, M., Miet, Maes, F., Schutyser, F., Marchal, G., Jacobs, R., & Suetens, P. (2006). Assessment of bone segmentation quality of cone-beam CT versus multislice spiral CT: a pilot study. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, 102(2), 225–234. <https://doi.org/10.1016/j.tripleo.2005.10.039>
- Luo, T., Shi, C., Zhao, X., Zhao, Y., & Xu, J. (2016). Automatic synthesis of panoramic radiographs from dental cone beam computed tomography data. *PLoS ONE*, 11(6). <https://doi.org/10.1371/journal.pone.0156976>
- Malladi, R., Sethian, J. A., & Vemuri, B. C. (1995). Shape Modeling with Front Propagation: A Level Set Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2), 158–175. <https://doi.org/10.1109/34.368173>
- Mårtensson, G., Ferreira, D., Granberg, T., Cavallin, L., Oppedal, K., Padovani, A., Rektorova, I., Bonanni, L., Pardini, M., Kramberger, M. G., Taylor, J.-P., Hort, J., Snædal, J., Kulisevsky, J., Blanc, F., Antonini, A., Mecocci, P., Vellas, B., Tsolaki, M., ... Westman, E. (2020). The reliability of a deep learning model in clinical out-of-distribution MRI data: a multicohort study. *Medical Image Analysis*, 66, 101714. <https://doi.org/10.1016/j.media.2020.101714>
- McRobbie, D. W., Moore, E. A., Graves, M. J., & Prince, M. R. (2006). MRI from picture to proton. In *MRI from Picture to Proton*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511545405>
- Militello, C., Rundo, L., & Gilardi, M. C. (2014). Applications of imaging processing to MRgFUS treatment for fibroids: a review. *Translational Cancer Research*, 3(5), 472–482. <https://doi.org/10.21037/3200>
- Militello, C., Vitabile, S., Rundo, L., Russo, G., Midiri, M., & Gilardi, M. C. (2015). A fully automatic 2D segmentation method for uterine fibroid in MRgFUS treatment evaluation. *Computers in Biology and Medicine*, 62, 277–292. <https://doi.org/10.1016/j.compbiomed.2015.04.030>
- Militello, C., Vitabile, S., Russo, G., Candiano, G., Gagliardo, C., Midiri, M., & Gilardi, M. C. (2013). A Semi-automatic Multi-seed Region-Growing Approach for Uterine Fibroids Segmentation in MRgFUS Treatment.

- Complex, Intelligent, and Software Intensive Systems (CISIS), 2013 Seventh International Conference On*, 176–182. <https://doi.org/10.1109/CISIS.2013.36>
- Minnema, J., Eijnatten, M., Hendriksen, A. A., Liberton, N., Pelt, D. M., Batenburg, K. J., Forouzanfar, T., & Wolff, J. (2019). Segmentation of dental cone-beam CT scans affected by metal artifacts using a mixed-scale dense convolutional neural network. *Medical Physics*, 46(11), 5027–5035. <https://doi.org/10.1002/mp.13793>
- Moon, T. K. (1996). The expectation-maximization algorithm. *Signal Processing Magazine, IEEE*, 13(6), 47–60. <https://doi.org/http://dx.doi.org10.1109/79.543975>
- Nardi, C., Borri, C., Regini, F., Calistri, L., Castellani, A., Lorini, C., & Colagrande, S. (2015). Metal and motion artifacts by cone beam computed tomography (CBCT) in dental and maxillofacial study. *La Radiologia Medica*, 120(7), 618–626. <https://doi.org/10.1007/s11547-015-0496-2>
- Ong, F., & Lustig, M. (2016). Beyond low rank + sparse: Multi-scale low rank matrix decomposition. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2016-May*, 4663–4667. <https://doi.org/10.1109/ICASSP.2016.7472561>
- Otsu, N. (1979). THRESHOLD SELECTION METHOD FROM GRAY-LEVEL HISTOGRAMS. *IEEE Trans Syst Man Cybern, SMC-9(1)*, 62–66. <https://doi.org/10.1109/tsmc.1979.4310076>
- Paatero, Y. V. (1961). Pantomography and orthopantomography. *Oral Surgery, Oral Medicine, Oral Pathology*, 14(8), 947–953. [https://doi.org/10.1016/0030-4220\(61\)90009-3](https://doi.org/10.1016/0030-4220(61)90009-3)
- Pal, N. R., Pal, K., Keller, J. M., & Bezdek, J. C. (2005). A possibilistic fuzzy c-means clustering algorithm. *IEEE Transactions on Fuzzy Systems*, 13(4), 517–530. <https://doi.org/10.1109/TFUZZ.2004.840099>
- Papakosta, T. K., Savva, A. D., Economopoulos, T. L., Matsopoulos, G. K., & Gröhdal, H. G. (2017). An automatic panoramic image reconstruction scheme from dental computed tomography images. *Dentomaxillofacial Radiology*, 46(4), 20160225. <https://doi.org/10.1259/dmfr.20160225>
- Pelt, D. M., & Sethian, J. A. (2017). A mixed-scale dense convolutional neural network for image analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 115(2), 254–259. <https://doi.org/10.1073/pnas.1715832114>
- Pianykh, O. S. (2006). Analytically tractable case of fuzzy c-means clustering. *Pattern Recognition*, 39(1), 35–46. <https://doi.org/10.1016/j.patcog.2005.06.005>
- Pizer, S. M., Johnston, R. E., Erickson, J. P., Yankaskas, B. C., & Muller, K. E. (1990). Contrast-limited adaptive histogram equalization: Speed and effectiveness. *Proceedings of the First Conference on Visualization in Biomedical*

- Computing*, 337–345. <https://doi.org/10.1109/vbc.1990.109340>
- Preda, L., Di Maggio, E. M., Dore, R., La Fianza, A., Solcia, M., Schifino, M. R., Campani, R., & Preda, E. G. (1997). Use of spiral computed tomography for multiplanar dental reconstruction. *Dento Maxillo Facial Radiology*, 26(6), 327–331. <https://doi.org/10.1038/sj.dmfr.4600290>
- Quinto, E. T. (2017). Artifacts and Visible Singularities in Limited Data X-Ray Tomography. *Sensing and Imaging*, 18(1), 1–14. <https://doi.org/10.1007/s11220-017-0158-7>
- Raghu, M., Zhang, C., Brain, G., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding Transfer Learning for Medical Imaging. *33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada*. <https://papers.nips.cc/paper/8596-transfusion-understanding-transfer-learning-for-medical-imaging.pdf>
- Reddy, T., & Kumaravel, N. (2010). Segmentation and Classification of Jaw Bone CT images using Curvelet based Texture features. *Bangladesh Journal of Medical Science*, 9(1), 33–43. <https://doi.org/10.3329/bjms.v9i1.5229>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
- Rousseau, F., Habas, P. A., & Studholme, C. (2011). A supervised patch-based approach for human brain labeling. *IEEE Transactions on Medical Imaging*, 30(10), 1852–1862. <https://doi.org/10.1109/TMI.2011.2156806>
- Rueckert, D., Frangi, A. F., & Schnabel, J. A. (2003). Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging*, 22(8), 1014–1025. <https://doi.org/10.1109/TMI.2003.815865>
- Rueda, S., Gil, J. A., Pichery, R., & Alcañiz, M. (2006). Automatic segmentation of jaw tissues in CT using active appearance models and semi-automatic landmarking. *Medical Image Computing and Computer-Assisted Intervention: MICCAI ...International Conference on Medical Image Computing and Computer-Assisted Intervention*, 9(1), 167–174. https://doi.org/10.1007/11866565_21
- Rundo, L., Militello, C., Tangherloni, A., Russo, G., Lagalla, R., Mauri, G., Gilardi, M. C., & Vitabile, S. (2019). Computer-assisted approaches for uterine fibroid segmentation in MRgFUS treatments: Quantitative evaluation and clinical feasibility analysis. In *Smart Innovation, Systems and Technologies* (Vol. 103, pp. 229–241). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-319-95095-2_22
- Rundo, L., Militello, C., Vitabile, S., Casarino, C., Russo, G., Midiri, M., & Gilardi, M. C. (2016). Combining split-and-merge and multi-seed region growing algorithms for uterine fibroid segmentation in MRgFUS treatments. *Medical*

- and Biological Engineering and Computing*, 54(7), 1071–1084.
<https://doi.org/10.1007/s11517-015-1404-6>
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- Sa-Ing, V., Wangkaom, K., & Thongvigitmanee, S. S. (2013). Automatic dental arch detection and panoramic image synthesis from CT images. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 6099–6102. <https://doi.org/10.1109/EMBC.2013.6610944>
- Salli, E., Aronen, H. J., Savolainen, S., Korvenoja, A., & Visa, A. (2001). Contextual clustering for analysis of functional MRI data. *Medical Imaging, IEEE Transactions On*, 20(5), 403–414. <https://doi.org/10.1109/42.925293>
- Schreiber, J. J., Anderson, P. A., Rosas, H. G., Buchholz, A. L., & Au, A. G. (2011). Hounsfield Units for Assessing Bone Mineral Density and Strength: A Tool for Osteoporosis Management. *The Journal of Bone and Joint Surgery-American Volume*, 93(11), 1057–1063. <https://doi.org/10.2106/JBJS.J.00160>
- Shannon, C. E. (1949). Communication in the Presence of Noise. *Proceedings of the IRE*, 37(1), 10–21. <https://doi.org/10.1109/JRPROC.1949.232969>
- Shen, D., & Davatzikos, C. (2000). An adaptive-focus deformable model using statistical and geometric information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 906–913. <https://doi.org/10.1109/34.868689>
- Shen, D., Wu, G., & Suk, H.-I. (2017). Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering*, 19, 221–248. <https://doi.org/10.1146/annurev-bioeng-071516-044442>
- Sørensen, T. (1948). *A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons*. 5(4), 1–34.
- Stein, W., Hassfeld, S., & Muhling, J. (1998). Tracing of thin tubular structures in computer tomographic data. *Computer Aided Surgery*, 3(2), 83–88. [https://doi.org/10.1002/\(SICI\)1097-0150\(1998\)3:2<83::AID-IGS5>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0150(1998)3:2<83::AID-IGS5>3.0.CO;2-G)
- Swennen, G. R. J., Mollemans, W., & Schutyser, F. (2009). Three-Dimensional Treatment Planning of Orthognathic Surgery in the Era of Virtual Imaging. *Journal of Oral and Maxillofacial Surgery*, 67(10), 2080–2092. <https://doi.org/10.1016/j.joms.2009.06.007>
- Székely, G., Kelemen, A., Brechbühler, C., & Gerig, G. (1996). Segmentation of 2-D and 3-D objects from MRI volume data using constrained elastic deformations of flexible Fourier contour and surface models. *Medical Image*

- Analysis*, 1(1), 19–34. [https://doi.org/10.1016/S1361-8415\(01\)80003-7](https://doi.org/10.1016/S1361-8415(01)80003-7)
- Tohka, J. (2003). Global optimization-based deformable meshes for surface extraction from medical images. In *Global optimization-based deformable meshes for surface extraction from medical images*. TTY.
- Tohnak, S., Mehnert, A., Crozier, S., & Mahoney, M. (2006). Synthesizing panoramic radiographs by unwrapping dental CT data. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, 3329–3332. <https://doi.org/10.1109/IEMBS.2006.260348>
- Torosdagli, N., Liberton, D. K., Verma, P., Sincan, M., Lee, J. S., & Bagci, U. (2019). Deep Geodesic Learning for Segmentation and Anatomical Landmarking. *IEEE Transactions on Medical Imaging*, 38(4), 919–931. <https://doi.org/10.1109/TMI.2018.2875814>
- Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 Bias Correction. *Medical Imaging, IEEE Transactions On*, 29(6), 1310–1320. <https://doi.org/http://dx.doi.org/10.1109/TMI.2010.2046908>
- Udupa, J. K., & Samarasekera, S. (1996). Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation. *Graphical Models and Image Processing*, 58(3), 246–261. <https://doi.org/10.1006/gmip.1996.0021>
- Umeyama, S. (1991). Least-Squares Estimation of Transformation Parameters Between Two Point Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4), 376–380. <https://doi.org/10.1109/34.88573>
- Unsalan, C., & Boyer, K. L. (2013). *Multispectral Satellite Image Understanding: From Land Classification to Building and Road Detection*. Springer Publishing Company, Incorporated.
- Us, D. (2019). *Reduction of Limited Angle Artifacts in Medical Tomography via Image Reconstruction* [Tampere University]. <https://trepo.tuni.fi//handle/10024/115305>
- Vaitiekūnas, M., Jegelevičius, D., Sakalauskas, A., & Grybauskas, S. (2019). Automatic Method for Bone Segmentation in Cone Beam Computed Tomography Data Set. *Applied Sciences*, 10(1), 236. <https://doi.org/10.3390/app10010236>
- Van Leemput, K., Maes, F., Vandermeulen, D., & Suetens, P. (1999). Automated model-based tissue classification of MR images of the brain. *Medical Imaging, IEEE Transactions On*, 18(10), 897–908. <https://doi.org/http://dx.doi.org/10.1109/42.811270>
- Vannier, M. W., Hildebolt, C. F., Conover, G., Knapp, R. H., Yokoyama-Crothers, N., & Wang, G. (1997). Three-dimensional dental imaging by spiral CT: A progress report. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontics*, 84(5), 561–570. [https://doi.org/10.1016/S1079-2104\(97\)90274-2](https://doi.org/10.1016/S1079-2104(97)90274-2)

- Viola, P., & Jones, M. (2001). *Rapid Object Detection using a Boosted Cascade of Simple Features*.
- Wang, L., Chen, K. C., Gao, Y., Shi, F., Liao, S., Li, G., Shen, S. G. F., Yan, J., Lee, P. K. M., Chow, B., Liu, N. X., Xia, J. J., & Shen, D. (2014). Automated bone segmentation from dental CBCT images using patch-based sparse representation and convex optimization. *Medical Physics*, 41(4), 043503. <https://doi.org/10.1118/1.4868455>
- Wang, L., Chen, K. C., Shi, F., Liao, S., Li, G., Gao, Y., Shen, S. G., Yan, J., Lee, P. K. M., Chow, B., Liu, N. X., Xia, J. J., & Shen, D. (2013). Automated Segmentation of CBCT Image Using Spiral CT Atlases and Convex Optimization. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013. MICCAI 2013. Lecture Notes in Computer Science*, 8151, 251–258. https://doi.org/10.1007/978-3-642-40760-4_32
- Wang, L., Gao, Y., Shi, F., Li, G., Chen, K. C., Tang, Z., Xia, J. J., & Shen, D. (2016). Automated segmentation of dental CBCT image with prior-guided sequential random forests. *Medical Physics*, 43(1), 336–346. <https://doi.org/10.1118/1.4938267>
- Worthington, P., Rubenstein, J., & Hatcher, D. C. (2010). The role of cone-beam computed tomography in the planning and placement of implants. *Journal of the American Dental Association*, 141(10 SUPPL.), 19S-24S. <https://doi.org/10.14219/jada.archive.2010.0358>
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2020). A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 1–21. <https://doi.org/10.1109/tnnls.2020.2978386>
- Yao, J., Chen, D., Lu, W., & Premkumar, A. (2006). Uterine fibroid segmentation and volume measurement on MRI. In A. Manduca & A. A. Amini (Eds.), *Medical Imaging 2006: Physiology, Function, and Structure from Medical Images* (Vol. 6143, p. 614322). SPIE. <https://doi.org/10.1117/12.653856>
- Yokota, F., Okada, T., Takao, M., Sugano, N., Tada, Y., Tomiyama, N., & Sato, Y. (2013). Automated CT segmentation of diseased hip using hierarchical and conditional statistical shape models. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8150 LNCS(PART 2), 190–197. https://doi.org/10.1007/978-3-642-40763-5_24
- Yun, Z., Yang, S., Huang, E., Zhao, L., Yang, W., & Feng, Q. (2019). Automatic reconstruction method for high-contrast panoramic image from dental cone-beam CT data. *Computer Methods and Programs in Biomedicine*, 175, 205–214. <https://doi.org/10.1016/j.cmpb.2019.04.024>
- Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability.

NeuroImage, 31(3), 1116–1128.
<https://doi.org/10.1016/J.NEUROIMAGE.2006.01.015>

Zhang, J., Liu, M., Wang, L., Chen, S., Yuan, P., Li, J., Shen, S. G. F., Tang, Z., Chen, K. C., Xia, J. J., & Shen, D. (2020). Context-guided fully convolutional networks for joint craniomaxillofacial bone segmentation and landmark digitization. *Medical Image Analysis*, 60, 101621. <https://doi.org/10.1016/j.media.2019.101621>

Zhang, J., Liu, M., Wang, L., Chen, S., Yuan, P., Li, J., Shen, S. G. F., Tang, Z., Chen, K. C., Xia, J. J., & Shen, D. (2017). Joint craniomaxillofacial bone segmentation and landmark digitization by context-guided fully convolutional networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10434 LNCS, 720–728. https://doi.org/10.1007/978-3-319-66185-8_81

PUBLICATIONS

PUBLICATION

I

Model-based segmentation of reconstructed dental x-ray volumes

Kari Antila, Mikko Lilja, Martti Kalke and Jyrki Lötjönen

Proceedings of the 2006 IEEE International Conference on Image Processing (ICIP), p. 1933-1936

<https://doi.org/10.1109/ICIP.2006.312792>

Publication reprinted with the permission of the copyright holders.

In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of Tampere University's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

MODEL-BASED SEGMENTATION OF RECONSTRUCTED DENTAL X-RAY VOLUMES

*Kari Antila**, *Mikko Lilja***, *Martti Kalke**** and *Jyrki Lötjönen**

*VTT Technical Research Centre of Finland, **Helsinki University of Technology, ***PaloDEx Group

ABSTRACT

Modern reconstruction algorithms allow volumetric imaging with conventional 2D dental X-ray systems. Volumetric images are useful in dental implantology, where the correct identification of key structures such as the edges of the mandible and the mandibular nerve is critical. This paper presents a segmentation method capable of extracting the mandible. The segmentation is based on a statistical model which was first transformed affinely and finally deformed non-rigidly to the object. The method was tested on three volumes with good results: mean distances between the deformed and manually segmented reference surfaces were 0.26, 0.34 and 0.50 mm. Applications of the method include the extraction of slices orthogonal to the mandibular bone centerline and local, anatomy based image enhancement.

Index Terms— Image Segmentation, Biomedical Image Processing, X-ray imaging

1. INTRODUCTION

The introduction of digital scanners for dental X-ray imaging has enabled totally new possibilities in acquisition, processing and visualization of the data. Furthermore, reconstruction techniques have been developed to handle incomplete imaging geometry and intensity information [1]. With such a technique, volumetric data (3D images) can be acquired using any existing digital X-ray system, including dental panoramic and intraoral systems. When compared to current techniques such as Computed Tomography (CT), overall costs and dose to the subject can be reduced substantially.

Unlike conventional 2D X-ray imaging, 3D imaging enables measuring distances and defining exact structures of the objects. For example in dental implantology, for the firm attachment of the implant it is critical to use the optimal screw size and therefore measure the optimal depth and angle for the screw hole. Upon failure to meet these criteria, the implant could be left loose or the mandibular nerve could be damaged causing possibly permanent facial numbness. Another example of the benefits of 3D imaging is the possibility to avoid the superposition problem, i.e. the problem caused by overlapping objects, in diagnostic imaging.

Model-based image segmentation has been studied extensively [2]. In this paper, we present a novel method for segmenting the mandible from reconstructed volumes. With the model used, prior information of typical shapes and relative positions of objects of interests can be exploited. The model can easily be supplemented with new objects like the mandible bone centerline, mental foramen and the mandibular nerve. As the algorithm should be applied to all images acquired during daily clinical practice, special attention was paid to the robustness and the computation time issues.

2. MATERIALS AND METHODS

2.1. Volumetric data and acquisition

In this work, image data to be segmented were acquired from three subjects with a scanner that generates images from a set of projection images (PaloDEx Group Oy, Tuusula, Finland). The prototype was configured to scan 11 projections with 3.3° spacing covering a limited angle of 33° . The scans were used in reconstructing isotropic, 12 bit grayscale volumes of size $256 \times 256 \times 256$ voxels with voxel size of $0.23 \times 0.23 \times 0.23 \text{ mm}^3$ [1]. Due to the limited size of the reconstruction volume, the mandibles were divided into three sections: right distal (the right cheek), mesial (the forward section of the jaw) and left distal.

2.2. Model construction

The average model of a mandible was constructed by manually segmenting the bone from a set of 31 rigidly aligned MR head volumes and taking the average of the manual segmentation results, as proposed in [3,4]. The average triangulated and shaded surface model is shown in Fig. 1.

As the whole mandible was not visible in the 3D reconstructions, only the visible section (left/right distal or mesial) of the average model was extracted (Fig. 1). This region-of-interest (ROI) was defined on the basis of the image acquisition protocol used. All phases of the segmentation were performed only on the extracted ROI model.

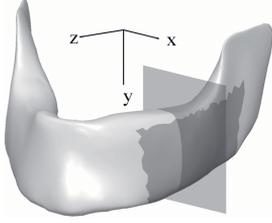


Fig. 1. The average mandible model surface. The ROI model, here the left distal section, extracted from the average model is shaded in dark gray. A typical coronal slice, (x,y) plane, (Fig. 2a) is marked with a dark square.

2.3. Affine transformation phase

The ROI model was first registered to the volume using an affine transformation. The optimal values for the rotation $\mathbf{R} = (r_x, r_y, r_z)$, translation $\mathbf{T} = (t_x, t_y, t_z)$ and scaling $\mathbf{S} = (s_x, s_y, s_z)$ parameters were searched by minimizing the energy function E_{affine} (read below Eq. 1).

The energy function was formulated to measure the goodness of fit between a binary volume and the ROI model. The binary threshold was selected heuristically by finding a grayscale value corresponding to 75 % percentile of all gray values in the volume. After thresholding, standard morphological operations, two erosions and two dilations, were applied (Fig. 2b). Then, oriented distance maps [5] were computed (Fig. 2c and d). If edges were close to each other in standard distance maps and the initialization of the model was not good, the distance map may have attracted the model to a wrong edge. This problem was partly solved by computing separate distance maps for differently oriented edges. In this work, six orientations (positive and negative x -, y - and z -axes) were used.

Each surface node was linked to one distance map by computing a dot product between the normal vector of the node and the orientation of the distance map, $D_{k(i)}$, and choosing the one producing the highest value. All normal vectors were pointing outwards from the surface. The mean distance of the nodes from the edges was computed from

$$E_{affine} = \frac{1}{N} \sum_i D_{k(i)}(\mathbf{p}_i), \quad (1)$$

where N is number of nodes in the model and \mathbf{p}_i denotes the location of node i linked to distance map k .

A global optimization technique was chosen to define the transformation parameters because the algorithm must be highly robust. However, global techniques [6], such as genetic algorithms or simulated annealing, are time consuming which would be a problem for our clinical tool. Therefore, we chose a special strategy for optimizing the parameters: the optimization was performed iteratively on a number of 2D cross-sectional slices of the volume and an

exhaustive search for a subset of parameters was done for each slice.

The optimization was done 1) on the coronal slice, (x, y) plane, at center of the image, such as the one in Fig. 2a, which provided an estimate of the subset $(r_z, t_x, t_y, s_x, s_y)$. Parameters (t_x, t_y) were applied to the model. The procedure was repeated 2) on the axial slice, (x, z) plane, in a position of 4.7 mm (20 voxels) above the bottom node of the model giving estimates of $(r_y, t_x, t_z, s_x, s_z)$. Out of these (r_y, t_z) were applied. In the final step 3) the matching was done on the same slice as in the first step. After the third step all parameters of the subset $(r_z, t_x, t_y, s_x, s_y)$ were applied. With these three steps an estimate of the subset $(r_y, r_z, t_x, t_y, t_z, s_x, s_y)$ was acquired. The rest (r_x, s_z) were considered constant. The parameters were searched through a space of $-8^\circ < r_y, r_z, < 8^\circ$ (with a spacing of 2°), $-20 < t_y, t_z < 20$ (voxels), $-50 < t_x < 50$ (voxels), (spacing 2 voxels) and $0.8 < s_x, s_y, s_z < 1.2$ (spacing 0.1).

2.4. Deformation phase

The final shape of the ROI model was searched with a non-rigid deformation method [5] based on a multi-resolution free-form deformation (FFD) grid. In this work, grid sizes up to $10 \times 10 \times 10$ were used. The points of the grid were moved to reach the minimum of the energy function

$$E_{deformation} = E_{data} + \gamma E_{model}, \quad (2)$$

where E_{data} is the mismatch energy between the model and the edges of the image, E_{model} retains undesired deformations of the model, and γ is a user-specified weight. In the original method, E_{data} is defined as in Eq. (1). Because thresholding does not separate the mandible accurately from the images, a novel energy measure was developed:

$$E_{data} = -\frac{1}{N} \sum_i g_i^-(g_i^- - g_i^+), \quad (3)$$

where

$$g_i^{+/-} = \frac{1}{m} \left(\sum_{j=1}^h I(\mathbf{p}_i + /- j\mathbf{n}_i) \right). \quad (4)$$

\mathbf{p}_i is the position of the model node i , \mathbf{n}_i is the surface normal of the node, h the user-defined length of the profile (here $h=4$), m is a user-specified constant (here $m=5$), and $I=I(x)$ is the gray value of the volume at x . In other words, the energy term drove the model points into locations where the contrast $(g^- - g^+)$ along surface normal of the model was maximized in the volume, and the gray-value was high (g^-). The energy component E_{model} regulated the change in orientation of the surface normals, and changes in the distances between the model nodes, described in detail in [5].

The whole segmentation method from the data and model input to the final result is summarized in. Fig. 3.

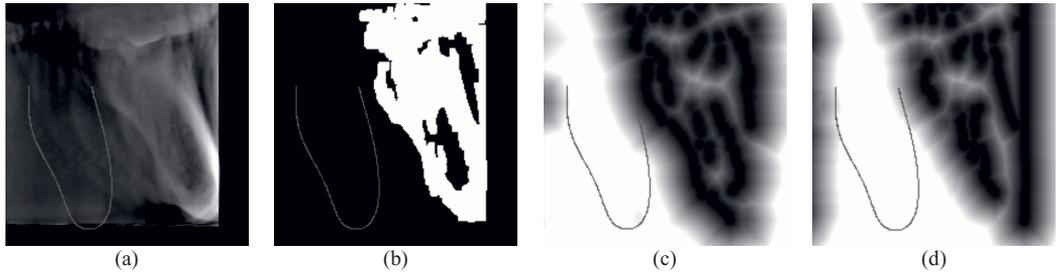


Fig. 2. (a) A coronal slice from a grayscale and (b) the corresponding binary volume taken from the left distal section. Distance maps oriented along (c) negative and (d) positive x-axis computed from the binary volume (b). A contour of the ROI model in its initial position is shown in dark outline.

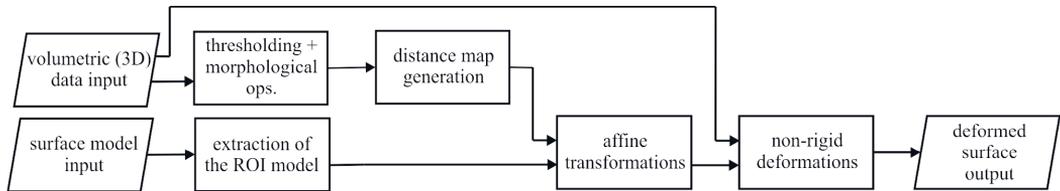


Fig. 3. A flowchart of the phases of the model-based segmentation method presented in section 2.

3. RESULTS

The tests were performed on three sample volumes (one right mesial, two left mesial). Example slices of the volumes with the initial model and the final surface after deformations are presented in Fig. 4. The goodness of the final fit was evaluated by segmenting the mandible surface manually and computing the distance of every node in the final surface to the manually segmented surface. The statistics are summarized in Table 1. All algorithms were implemented in the C programming language. The total computation time was approximately 18 s on an Intel Pentium 4 class workstation.

Table 1. The results of computing the distances (mm) from the manually segmented surfaces to surfaces after affine transformation (aff) and deformation (nrg) phases.

	mean (aff)	mean (nrg)	std. (aff)	std. (nrg)	RMS (aff)	RMS (nrg)
1	0.4209	0.2614	0.0150	0.0099	0.5330	0.3393
2	0.7980	0.3387	0.0361	0.0135	1.0849	0.4361
3	0.9001	0.4995	0.0350	0.0173	1.1589	0.6159

The results in Table 1 show that the non-rigid deformation phase improved the accuracy significantly, especially when the sample shape was clearly distinct from the ROI model shape which was the case with samples 2 and 3.

As mentioned in the introduction, the segmentation of the mandible is necessary for several reasons, such as

computing slices orthogonal to the mandible, needed e.g. for better detecting the mandibular nerve. First, the centerline of the average model was determined manually. Thereafter, it was deformed to match the volumes along with the ROI model, according to the transformation defined during the segmentation. Finally, the slices orthogonal to the mandible were computed from the volume using trilinear 3D interpolation [7]. An example of an axial slice with the registered centerline and a computed orthogonal slice are presented in Fig. 5.

4. DISCUSSION AND CONCLUSIONS

The benefit of using a set of 2D transformation parameter optimizations instead of 3D optimization was significantly better computational efficiency. In 2D, a large number of parameter combinations were tested fast, and the risk of converging to a local minimum was avoided. The computational performance of the affine transformation phase could be improved further by developing sequential search methods based on multi-resolution and adopting a global optimization in 2D, such as [6]. Currently the computation time is linearly dependent on size of the searched parameter domain.

The segmentation method produced accurate results reasonably fast. However, as the method should be highly robust, the validation of the algorithm with much higher number of cases must be performed as more images become available. Future releases will include new features such as the extraction of the mandibular nerve shape and position.

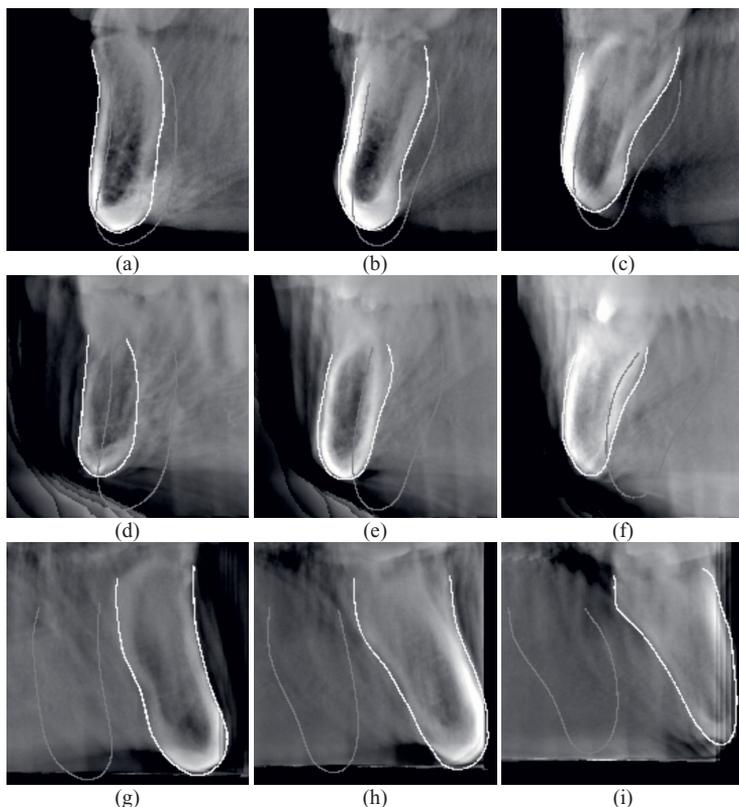


Fig. 4. The results of the segmentation. Three example slices (in rows) were selected from three data samples (a,b,c), (d,e,f) and (g,h,i). Initial position of the ROI model is marked with a dark outline while the final surface is marked white.

6. REFERENCES

- [1] R. Gordon, R. Bender and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography," *J. Theoret. Biol.*, Vol. 29, pp. 471-482, 1970.
- [2] C. Xu, D. Pham and J. Prince, "Image Segmentation Using Deformable Models," *Handbook of Medical Imaging, Volume 2. Medical Imaging, Processing and Analysis*, pp. 129-174 SPIE Press, Bellingham, Washington, USA, 2000.
- [3] A. Guimond, J. Meunier, and J-P. Thirion, "Average Brain Models. A Convergence Study," *Computer Vision and Image Understanding*, Vol. 77, No. 2, pp. 192-210, 2000.
- [4] D. Rueckert, A. Frangi, and J. Schnabel, "Automatic Construction of 3-D Statistical Deformation Models of the Brain Using Nonrigid Registration," *IEEE Transactions on Medical Imaging*, Vol. 22, No. 8, pp. 1014-1025, 2003.
- [5] J. Lötjönen, P-J. Reissman, I. E. Magnin, and T. Katila, "Model extraction from magnetic resonance volume data using the deformable pyramid," *Medical Image Analysis*, Vol. 3, No. 4, pp. 387-406, 1999.
- [6] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, "Lipschitzian optimization without the lipschitz constant," *Journal of Optimization Theory and Applications*, Vol. 79, pp.157-181, 1993.
- [7] J. A. Parker, R. V. Kenyon, and D. E. Troxel, "Comparison of interpolating methods for image resampling," *IEEE Transactions on Medical Imaging*, Vol. MI-2, pp. 31-39, 1983.

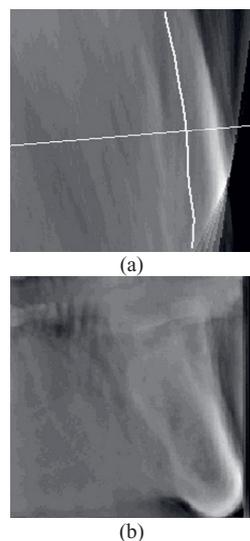


Fig. 5. (a) An axial slice of a volume and (b) the orthogonal slice corresponding to the middle point of the registered bone centerline. In (a) the centerline is the thick vertical line and the thin horizontal line orthogonal to it illustrates the orientation of the orthogonal slice.

PUBLICATION II

Automatic segmentation of the mandible from limited-angle dental x-ray tomography reconstructions

Mikko Lilja, Ville Vuorio, Kari Antila, Henri Setälä, Jorma Järnstedt and Mika Pollari

Proceedings of the 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), p. 967-967

<https://doi.org/10.1109/ISBI.2007.357014>

Publication reprinted with the permission of the copyright holders.

In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of Tampere University's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

AUTOMATIC SEGMENTATION OF THE MANDIBLE FROM LIMITED-ANGLE DENTAL X-RAY TOMOGRAPHY RECONSTRUCTIONS

Mikko Lilja^{1,*}, Ville Vuorio^{1,*}, Kari Antila^{2,*}, Henri Setälä³, Jorma Järnstedt⁴, and Mika Pollari^{1,*}

¹ Helsinki University of Technology, Laboratory of Biomedical Engineering, P. O. Box 2200, FI-02015 HUT, Finland

² VTT Technical Research Centre, VTT Information Technology, P. O. Box 1206, FI-33101 Tampere, Finland

³ PaloDEX Group Oy, Tykistökatu 4 B, FI-20520 Turku, Finland

⁴ Pirkanmaa Hospital District, Regional Medical Imaging Centre, P. O. Box 2000, FI-33521 Tampere, Finland

ABSTRACT

A 3-D reconstruction from sparse limited-angle x-ray projection data is a useful compromise between a single radiograph and a full CT reconstruction, e.g. in dental imaging. The segmentation of such volumes is desirable for clinical applications such as implantology, but the task is complicated by the inherent limited spatial validity of the reconstructions. We present an automatic model-based method for extracting the mandible from 3-D limited-angle dental x-ray reconstructions. The process includes enhancing the reconstruction, estimating the successfully reconstructed mandibular area, and the actual segmentation process. The results with 13 reconstructions are good with an average segmentation error of 0.32 mm.

Keywords: Image segmentation, X-ray tomography

1. INTRODUCTION

The reconstruction of limited-angle projection data is an ill-posed problem and prone to errors. Despite its shortcomings, limited-angle tomography (LAT) has interesting applications, e.g., in dental x-ray imaging as a low-dose and low-cost 3-D imaging modality [1]. One of the applications benefiting from a 3-D image is implantology [2].

To make the most of a dental LAT reconstruction, extracting the mandible is essential: it enables specific image enhancement within the mandible to improve the identification of internal features such as the mandibular canal. However, the scarce projection data causes artifacts and limits the spatial validity of a LAT reconstruction as illustrated by Fig. 1(b)—the reconstruction is not valid in the incisor (top) region, where the mandible curvature exceeds the angle of view (AOV). This complicates the segmentation of such reconstructions.

Deformable models have been successfully used in several segmentation applications, but such methods cannot cope with the limited spatial validity of LAT reconstructions. As one solution, the direct use of limited-angle sinogram data for segmentation has been suggested [3]. We propose the use of a standard global deformable model adapted to the local LAT reconstruction based on the model shape and the projection angles.

In this paper we present a method for extracting the mandible from dental x-ray LAT reconstructions. In Sect. 2.3 we introduce our

*The research was supported by TEKES, the National Technology Agency of Finland.

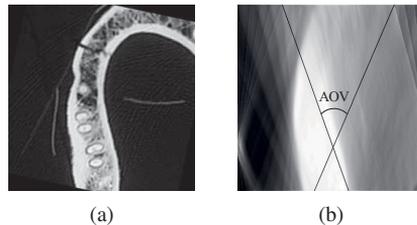


Fig. 1. An example of the limited validity of a LAT reconstruction: (a) is an axial cross section of a CT volume of a dry mandible and (b) is the LAT equivalent with the AOV overlaid in black.

mandible model, and based on it, in Sect. 2.4 propose a method for estimating the valid region of the reconstruction suitable for segmentation. Then, in Sect. 2.5 we present a filtering process for reducing artifacts and improving contrast in the reconstruction. Finally, in Sects. 2.6–3 we present an automatic segmentation process and the results of applying it on a set of 13 reconstructions.

2. MATERIAL AND METHODS

2.1. Limited-angle tomography reconstructions

Our image material consisted of 13 LAT reconstructions courtesy of PaloDEX Group Oy (Tuusula, Finland). Nine of these represented dry human mandibles (bone and teeth, no other tissues) and four represented volunteer patients. For each reconstruction 11 projections were acquired at an interval of approximately $4\text{--}5^\circ$ using an Instrumentarium Orthopantomograph® OP200 D dental x-ray device. A bite plate with 5 embedded pellets was used for patient positioning. The pellets were cast in a U-shape and served as radiopaque landmarks in the projections. A 16-bit $256 \times 256 \times 256$ gray scale volume was reconstructed from the projections using a modified algebraic reconstruction technique (ART). The volumes covered a $60 \times 60 \times 60$ mm³ mandibular region (voxel size 0.23 mm). Six imaging sections were used to cover the entire mandible corresponding to left and right incisor, premolar, and molar regions.

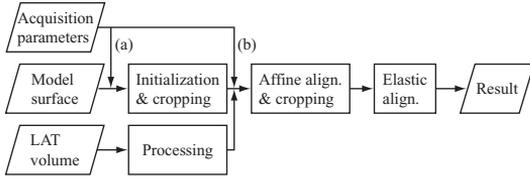


Fig. 2. A flowchart of the segmentation process. Parameters (a) are the 3-D coordinates of the pellets and the mandible section, and (b) the projection angles.

2.2. Overview of the segmentation process

Our segmentation method is presented as a flowchart in Fig. 2 and briefly described here. Before the segmentation, we perform a filtering on the reconstruction according to Sect. 2.5 to reduce artifacts and improve contrast. Then we use the landmark pellets to roughly register our mandible surface model to the reconstruction, and based on the mandible section (right/left, incisor/premolar/molar) we crop our model to crudely match the region of interest. Then we perform an affine alignment process during which we refine the estimate of the validly reconstructed region using the projection angles, as described in Sects. 2.4 and 2.7. Finally, we perform an elastic transformation in two stages according to Sect. 2.8. The result is a surface segment estimating the validly reconstructed region of the mandibular outline.

The segmentation method was implemented in MATLAB. The parameters of the transformations were optimized in bounded search spaces using simplex optimization to minimize the energy measure presented in Sect. 2.6.

2.3. Mandible model

To construct a mandible surface model, the mandible was manually extracted from 9 dental CT volumes provided by Pirkanmaa Hospital District (Tampere, Finland), and 31 MR volumes of the head. The surfaces were rigidly aligned according to [4], and their average was taken to create a model surface Ω with 7407 nodes.

The axial centerline γ of Ω was estimated by projecting its nodes to the xy plane and fitting a parabola to the point set corresponding to the dental region. The bite plate pellets were manually located from a CT volume of a dry mandible attached to a bite plate, and then rigidly added to the model. The surface Ω was divided in 3 sections: inferior lateral (Ω_A), inferior medial (Ω_B), and superior (Ω_C) as shown in Fig. 3. In the reconstructions the mandible generally exhibited the best contrast in Ω_A and the worst in Ω_C . This was taken into account in computing the energy measure and in choosing the region of interest in the affine transformations.

2.4. Estimating the region of valid reconstruction

One of the main difficulties related to the segmentation of limited-angle reconstructions is that only a fraction of the volumetric data is valid, i.e., correctly describes the geometry of the object. Areas, where the reconstruction fails, are not suitable for model-based segmentation. If these are not excluded, nonexistent features in the

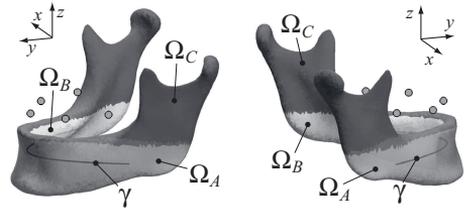


Fig. 3. The mandible model Ω (7407 nodes) and the marker pellets as seen from front and back. The mandibular centerline γ and the division into three sections Ω_A (inferior lateral), Ω_B (inferior medial), and Ω_C (superior) are illustrated.

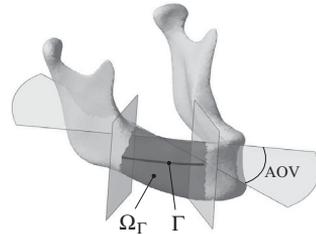


Fig. 4. An example of cropping the model: the mandible surface is cut perpendicular to Γ at its end points. The resulting surface segment Ω_Γ is rendered in dark gray.

reconstruction may direct the model to a false result. As shown by Quinto [5], only edges, which are tangential to some line of projection, can be reliably reconstructed from limited-angle projection data. Based on this, we use the tangential directions of the mandible centerline γ to estimate, which part $\Omega_\Gamma \subset \Omega$ of the mandible has been adequately reconstructed in the volume and is suitable for segmentation.

Because the initial position of the model relative to the projection directions may be inaccurate, we apply the cropping process described below only after a few iterations of affine alignment. The initial estimate of the successfully reconstructed region is based on the mandibular section (right/left, incisor/premolar/molar) and prior knowledge, and denoted $\Omega_{\Gamma,0}$.

After the first iterations of affine alignment, we determine the centerline segment Γ , whose tangential directions fall within the AOV spanned by the projections. Then we extract $\Omega_\Gamma \subset \Omega$ as the model segment, which remains between the two planes intersecting Γ perpendicularly at its endpoints. This is illustrated in Fig. 4. The final region of interest (ROI) is then defined $\hat{\Omega} = \Omega_{\Gamma,0} \cap \Omega_\Gamma$, i.e., at most the estimated valid region Ω_Γ is matched to the reconstruction.

2.5. Pre-processing of the reconstructions

Due to low contrast and artifacts present in LAT reconstructions, the volumes are pre-processed prior to the segmentation. The goal of pre-processing is to enhance the contrast of the mandible outline and to decrease the intensity of the artifacts.

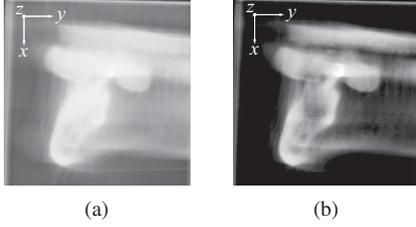


Fig. 5. The effect of pre-processing: (a) is a coronal cross section from a molar LAT reconstruction and (b) the same image after processing.

Volumes consist of clearly separable brighter and darker regions. The area of interest is located in the brighter region with some artifacts, while the darker region contains only artifacts. We divide the volume into bright and dark regions by using watershed segmentation. The dynamic range of the volume is expanded with parameters determined from the histogram of the bright region. Intensity values between 1–99.9% of the histogram mass are expanded to the full dynamic range of the volume. This weakens the artifacts in the darker region while increasing the contrast of the brighter region.

Finally, we process the brighter region locally. Local processing is used because the mandible edge has a locally higher intensity than the artifacts, but not necessarily globally due to the intensity variation in the volume. The brighter region is processed in $32 \times 32 \times 32$ voxel blocks. The dynamic range of each block is expanded by moving the black point to zero. An example of the processing result is presented in Fig. 5.

2.6. Energy measure

To describe how well a region of interest $\tilde{\Omega} \subset \Omega$ matches the mandible’s outline in the volume V , we use a slightly modified version of the energy measure we proposed in [6], namely

$$E_{\text{data}}(\tilde{\Omega}) = \frac{1}{\mu(\tilde{\Omega})} \int_{\tilde{\Omega}} V(\mathbf{r}) \mathbf{n}(\mathbf{r}) \cdot \nabla V(\mathbf{r}) d\Omega \\ \approx \frac{1}{2Nh} \sum_{i=1}^N V(\mathbf{x}_i) [V(\mathbf{x}_i + h\mathbf{n}_i) - V(\mathbf{x}_i - h\mathbf{n}_i)], \quad (1)$$

where $V(\mathbf{x}_i)$ is the gray scale value and \mathbf{n}_i is the outward unit surface normal vector of Ω at $\mathbf{x}_i = [x_i \ y_i \ z_i]$, N is the number of nodes in $\tilde{\Omega}$ and μ stands for area. We compute the mean difference value using $h = 1, \dots, 4$ for robustness.

The measure (1) is minimized, when $\tilde{\Omega}$ surrounds a bright area, which corresponds to the general appearance of the mandible in an x-ray image. The section Ω_C is an exception, as teeth may appear brighter than the mandible (e.g., in case P4 of Fig. 6) and therefore in $\tilde{\Omega} \cap \Omega_C$ the absolute value of the normal derivative is multiplied by -1 .

In a global affine transformation $E_{\text{total}} = E_{\text{data}}$. To constrain the elastic transformations as proposed in [7, 8], an internal energy measure of the form

$$E_{\text{model}}(\tilde{\Omega}) = \frac{1}{N} \sum_{i=1}^N w_1 \mathbf{n}_i \cdot \mathbf{n}_{i,0} + w_2 \|\mathbf{r}_i - \mathbf{r}_{i,0}\|^2 \quad (2)$$

was used to compute $E_{\text{total}} = E_{\text{data}} + E_{\text{model}}$. In (2) $\mathbf{n}_{i,0}$ is the reference direction of the surface normal and $\mathbf{r}_{i,0}$ the reference location of the i^{th} node and $w_1 \leq 0$, $w_2 \geq 0$ are suitable weights.

2.7. Model initialization and affine alignment

The five landmark pellets are used to initially register the model to the volume. The 2-D locations of the pellets are detected from the projections with a template matching method, and these are solved for the 3-D coordinates in the reconstruction. The mandible model is then initialized to the reconstruction by using the rigid transformation [4] between the pellet locations.

The degrees of freedom (DOFs) of a 3-D affine transformation are translation $\mathbf{t} = [t_x \ t_y \ t_z]$, rotation $\mathbf{r} = [r_x \ r_y \ r_z]$, scaling—either isotropic s_{xyz} or anisotropic $\mathbf{s} = [s_x \ s_y \ s_z]$ —and shear $\boldsymbol{\sigma} = [\sigma_{xy} \ \sigma_{xz} \ \sigma_{yz}]$. The model is first deformed using different combinations of these DOFs and different ROIs for computing the energy measure (1). Presented as tuples (DOFs, ROI), the stages of the affine transformation are $(\{\mathbf{t}, \mathbf{r}\}, \Omega_{\Gamma,0} \cap \Omega_A)$, $(\{\mathbf{t}, \mathbf{r}, s_{xyz}\}, \Omega_{\Gamma,0})$, and $(\{\mathbf{t}, \mathbf{r}, \mathbf{s}, \boldsymbol{\sigma}\}, \Omega_{\Gamma,0} \cap \Omega_{\Gamma})$. Each stage is repeated until the energy value changes less than 1% in one iteration.

For example in molar cases the search spaces of the DOFs are $\Delta \mathbf{t} = \pm[20 \ 30 \ 10]$ mm, $\Delta \mathbf{r} = \pm[15 \ 15 \ 35]^\circ$, $0.75 \leq s \leq 1.33$, and $\Delta \boldsymbol{\sigma} = \pm[0.5 \ 0.5 \ 0.5]$. The coordinates refer to the reconstructions as illustrated in Fig. 5. The final values of the nodal normal directions and locations are taken as the references $\mathbf{n}_{i,0}$, $\mathbf{r}_{i,0}$ for computing E_{model} in the subsequent elastic transformations.

2.8. Elastic alignment

After the affine transformations, the mandibular surface is deformed elastically. First, $2 \times 2 \times 2$ and $3 \times 3 \times 3$ FFD grids and trilinear interpolation are used to deform $\tilde{\Omega} = \Omega_{\Gamma,0} \cap \Omega_{\Gamma}$. The search spaces for the FFD control points are 33% of the grid’s respective dimensions. Then, $\tilde{\Omega}$ is deformed locally by moving its nodes in the direction of the surface normals and distributing the displacements to the rest of nodes in Ω in a smooth Gaussian manner, similarly to the method presented in [8]. The displacements are computed according to

$$\mathbf{d}_k = \exp\left(-\frac{\|\mathbf{r}_j - \mathbf{r}_k\|^2}{2\sigma^2}\right) d \mathbf{n}_j, \quad (3)$$

where $\mathbf{r}_j \in \tilde{\Omega}$ is the location of the displaced node, \mathbf{n}_j its outward unit surface normal, d its displacement, $\mathbf{r}_k \in \Omega$ is an arbitrary node location, and \mathbf{d}_k its displacement. Values 4.5 mm and 3.5 mm are used for σ , and 33% of the nodes in $\tilde{\Omega}$ are uniformly sampled for optimization in both cases. The search spaces for the normal displacement are equal to the parameter σ .

3. RESULTS

The segmentation process was performed on 13 reconstructions. On a 2.4 GHz Pentium 4 workstation with 1 GB of memory, the segmentation of one reconstruction took approximately 3–10 minutes, of which 75% was due to the local Gaussian deformation. The results were evaluated by calculating the average shortest Euclidean distance of $\tilde{\Omega}$ to the manually extracted mandible surface, which was ensured to exceed the coverage of $\tilde{\Omega}$.

Table 1. Mean values and standard deviations of the segmentation error (in mm) at different stages. Cases D1–D9 refer to volumes acquired from a dry mandible and P1–P4 to volumes acquired from patients. The first letter of the abbreviated mandible section stands for right/left and the second for incisor/premolar/molar.

Case	Sect.	AOV	Affine mean \pm SD	FFD mean \pm SD	Gaussian mean \pm SD
D1	LI	41.4°	0.39 \pm 0.65	0.26 \pm 0.48	0.17 \pm 0.36
D2	RP	42.7°	0.44 \pm 0.72	0.22 \pm 0.45	0.26 \pm 0.55
D3	RM	42.4°	0.37 \pm 0.63	0.26 \pm 0.53	0.14 \pm 0.39
D4	LM	50.4°	0.91 \pm 1.01	0.61 \pm 0.89	0.48 \pm 0.68
D5	LI	49.9°	0.42 \pm 0.69	0.25 \pm 0.55	0.21 \pm 0.52
D6	LP	49.7°	0.67 \pm 1.02	0.46 \pm 0.79	0.23 \pm 0.51
D7	LM	50.4°	1.01 \pm 1.28	0.55 \pm 0.84	0.56 \pm 1.06
D8	RM	50.9°	0.59 \pm 1.05	0.55 \pm 0.96	0.44 \pm 0.91
D9	RP	50.3°	0.20 \pm 0.51	0.29 \pm 0.67	0.21 \pm 0.55
	mean		0.55 \pm 0.84	0.38 \pm 0.68	0.30 \pm 0.61
P1	LM	39.2°	0.52 \pm 0.70	0.25 \pm 0.48	0.16 \pm 0.37
P2	RM	38.4°	1.03 \pm 1.42	0.88 \pm 1.29	0.71 \pm 1.20
P3	LM	38.1°	0.68 \pm 0.93	0.61 \pm 0.92	0.35 \pm 0.73
P4	RP	42.9°	0.43 \pm 0.77	0.33 \pm 0.67	0.29 \pm 0.58
	mean		0.66 \pm 0.95	0.52 \pm 0.84	0.38 \pm 0.72
	mean total		0.59 \pm 0.87	0.42 \pm 0.73	0.32 \pm 0.65

The segmentation errors are presented in Table 1 and six examples of the initial and converged mandible outlines are presented in Fig. 6.

4. DISCUSSION AND CONCLUSIONS

As can be seen from the results, the segmentation error generally decreases with each stage. The final error ranges from 0.14–0.56 mm for the dry mandibles and 0.16–0.71 mm for the patients. The mean value is 0.32 mm, i.e., less than 1.5 voxels. The result with dry mandibles is slightly better than with patients, but the difference can be attributed to case P2 (see Fig. 6), which exhibits poor contrast even after pre-processing. The processing improved the reconstructions and enabled a better segmentation result. In some cases it was in fact necessary for the model to converge to the correct location.

The initial rigid registration using the landmark pellets served its purpose well in the sense that in each case the model converged to a visually plausible result starting from the initialization. However, the pellets were added to the mandible model based on a single CT reconstruction, and the variability of the bite plate location in a patient’s mouth was not studied.

Our approach for estimating the region of interest proved to give reasonable results as the final ROI $\hat{\Omega}$ generally converged to the area, which appeared correctly reconstructed. It provided an automatic way of cropping the mandible model for extracting a small region of a local reconstruction, and enabled optimizing the transformations based on anatomically valid intensity information.

The segmentation results were good, but on the other hand the parameters and search spaces were chosen by trial and error, so that the segmentation succeeded in each case. However, these parameters

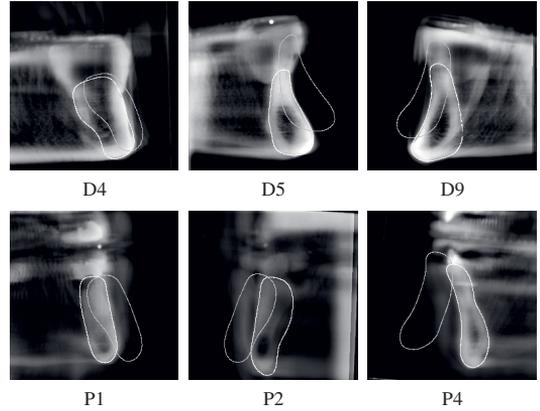


Fig. 6. Examples of the segmentation results overlaid on cross sections from the pre-processed reconstructions. The titles correspond to the cases in Table 1. The initial mandibular outline corresponding to the marker pellet locations is overlaid in light gray and the converged result in white.

were the same for each case. As more data becomes available, the segmentation method can be more thoroughly evaluated.

5. REFERENCES

- [1] R.L. Webber et al., “Tuned-aperture computed tomography (tact): theory and application for three-dimensional dentoalveolar imaging,” *Dentomaxillofac. Radiol.*, vol. 26, pp. 53–62, 1997.
- [2] A. Ekestubbe et al., “The use of tomography for dental implant planning,” *Dentomaxillofac. Radiol.*, vol. 26, pp. 206–213, 1997.
- [3] V. Elangovan and R.T. Whitaker, “From sinograms to surfaces: A direct approach to the segmentation of tomographic data,” in *Proc. MICCAI 2001*, 2001, pp. 213–223.
- [4] S. Umeyama, “Least-squares estimation of transformation parameters between two point patterns,” *Med. Imag. Anal.*, vol. 13, pp. 376–380, 1991.
- [5] E.T. Quinto, “Singularities of the x-ray transform and limited data tomography in \mathbb{R}^2 and \mathbb{R}^3 ,” *SIAM J. Math. Anal.*, vol. 24, pp. 1215–1225, 1993.
- [6] K. Antila et al., “Model-based segmentation of reconstructed dental x-ray volumes,” in *Proc. ICIP 2006*, 2006, pp. 1933–1936.
- [7] J. Lötjönen et al., “Model extraction from magnetic resonance volume data using the deformable pyramid,” *Med. Imag. Anal.*, vol. 3, pp. 387–406, 1999.
- [8] J. Lötjönen and T. Mäkelä, “Elastic matching using a deformation sphere,” in *Proc. MICCAI 2001*, 2001, pp. 541–548.

PUBLICATION III

**Automatic extraction of mandibular bone geometry for anatomy-based
synthetization of radiographs**

Kari Antila, Mikko Lilja, Martti Kalke and Jyrki Lötjönen

Proceedings of the 30th Annual International Conference of the IEEE Engineering in
Medicine and Biology Society (EMBC) 2008
<https://doi.org/10.1109/IEMBS.2008.4649197>

Publication reprinted with the permission of the copyright holders.

In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of Tampere University's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

Automatic Extraction of Mandibular Bone Geometry for Anatomy-based Synthetization of Radiographs

Kari Antila, Mikko Lilja, Martti Kalke and Jyrki Lötjönen, *Member, IEEE*

Abstract—We present an automatic method for segmenting Cone-Beam Computerized Tomography (CBCT) volumes and synthetizing orthopantomographic, anatomically aligned views of the mandibular bone. The model-based segmentation method was developed having the characteristics of dental CBCT, severe metal artefacts, relatively high noise and high variability of the mandibular bone shape, in mind. First, we applied the segmentation method to delineate the bone. Second, we aligned a model resembling the geometry of orthopantomographic imaging according to the segmented surface. Third, we estimated the tooth orientations based on the local shape of the segmented surface. These results were used in determining the geometry of the synthetized radiograph. Segmentation was done with excellent results: with 14 samples we reached 0.57 ± 0.16 mm mean distance from hand drawn reference. The estimation of tooth orientations was accurate with error of 0.65 ± 0.80 degrees. An example of these results used in synthetizing panoramic radiographs is presented.

I. INTRODUCTION

Applications in dental and maxillofacial surgery require accurate knowledge of the structure and geometry of the area to be operated. Popular planar tomography method, orthopantomography, delivers radiographs of very high resolution and wide field of view with small cost, great ease of use and low patient dose. These advantages have made the modality widely accepted and used. Its weaknesses, however, are superposition, blurring and loss of depth information, in short, incapability to capture volumetric information. The demand for volumetric information with advances in image reconstruction techniques has made Cone-Beam Computerized Tomography (CBCT) [1] and variants popular. CBCT delivers diagnostic value equivalent to conventional Computerized Tomography (CT) [2] but with only somewhat higher dose and cost compared to orthopantomography.

Availability of volumetric images potentially eliminates a great deal of the guesswork in planning of dental and

maxillofacial surgeries. However, finding the right structure, view and slice orientation may still require considerable amount of manual work and give inaccurate and non-reproducible results. Therefore, automated image processing methods are needed to make the analysis of images more straightforward and reliable.

Our objective is to develop tools for a dentist to extract needed information from CBCT images. This objective is reached by 1) *developing an automatic model-based segmentation method* capable of accurate mandibular bone delineation from CBCT volumes, 2) *extracting necessary measures of bone shape* from the segmentation result and by 3) *developing necessary tools for synthetizing optimized panoramic radiographs* based on the results of steps 1-2.

Practically every CT manufacturer [3]–[5] and OEM vendors [6] have developed workstation software for dental applications. Methods for mandibular segmentation from CT images have been proposed [7],[8]. However accurate mandibular segmentation of CBCT data still remains a challenge [9] as it must tolerate the characteristics of typical dental CBCT volumes like relatively high level of noise [10], severe metal artefacts and present or missing teeth. Image segmentation is one of the major challenges in image processing also in general. However, it is a pre-requisite for many image processing and analysis operations.

The goal of our approach was to develop a method which maintains the useful features of orthopantomographic, or shortly panoramic radiographs and overcome their main weakness, the loss of depth information. In this work, we developed tools allowing the extraction of such radiographs from CBCT volumes.

II. MATERIALS AND METHODS

We validated the methods by using 14 CBCT volumes, acquired with Soredex Scanora 3D scanner. The volumes included data from the center, right and left of the mandible spanning either a narrow or a wide cylinder-shaped reconstructed area of the mandible. Many of the reconstructions contained severe artifacts caused by metal fillings and implants. These heterogeneous volumes were used in order to evaluate robustness in clinical practice. The dimensions of the volumes varied between $300 \times 300 \times 288$ and $450 \times 450 \times 450$ voxels, and the voxel dimensions were between $0.13 \times 0.13 \times 0.13$ and $0.35 \times 0.35 \times 0.35$ mm.

Data used in the estimation of tooth orientations were collected by manually identifying and indexing all visible

Manuscript received April 16, 2008. This work was supported by TEKES, the National Technology Agency of Finland. K. Antila was supported by the Finnish Cultural Foundation and the Instrumentarium Science Foundation.

K. Antila is with Laboratory of Mathematics in Imaging, Brigham and Women's Hospital, Harvard Medical School and VTT Technical Research Centre of Finland, P.O. Box 1300, FI-33101 Tampere, Finland (phone: +358 40 834 7509, e-mail: kari.antila@vtt.fi).

M. Lilja is with Helsinki University of Technology, P.O. Box 1000, FI-02015 TKK, Finland.

M. Kalke is with PaloDEX Group Inc., P.O. Box 64, FI-04301 Tuusula, Finland.

J. Lötjönen is with VTT Technical Research Centre of Finland.

teeth in the data measuring the respective angle of tooth orientation in a plane orthogonal to the bone. The library consisted of 112 tooth locations and angles of orientation.

A. The Mandibular and Orthopantomogram Model

A mean statistical mandible model surface \mathcal{S} (see Fig. 1) with 7407 nodes/14810 triangles was created by manually determining the mandibular bone outline from 9 dental CT reconstructions of the mandible and 31 Magnetic Resonance (MR) images of the head. The average of the rigidly aligned surfaces was recomputed iteratively by finding the initial average surface, fitting it to the binary masks of the manually drawn references and computing a new average. This was repeated until a satisfactory point correspondence was reached.

We have previously used the same model for segmenting limited-angle dental x-ray reconstructions [11]. As previously, a parabolic approximation was used for the 3D centerline of the surface. Furthermore, the centerline, denoted C , was now used to fix the coordinate system of the model, i.e., the set $\{x_1, x_2, x_3, o\}$, where the vectors x_i constitute an orthonormal basis and o is the origin in Fig. 1. This model coordinate system is used for computing the affine transformations described in Section II B. This approach corresponds to the Lagrangian description of motion with a moving coordinate system as opposed to the fixed laboratory coordinate system in the Eulerian framework. The benefit of fixing the frame of transformation with the model is that the linear degrees of freedom (DOFs) always affect the same property of the model regardless of its attitude w.r.t. the CBCT coordinate system—e.g., x_1 scaling affects solely the width of the model.

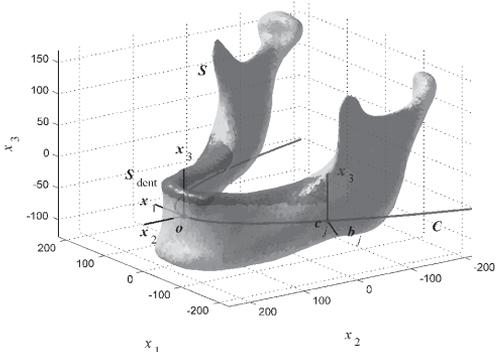


Fig. 1. The centerline C was fitted to the mandibular bone mean model \mathcal{S} . The model coordinate system $\{x_1, x_2, x_3, o\}$, the basis vectors of a normal plane of the centerline, and the dental area $\mathcal{S}_{\text{dent}}$ have also been illustrated. Vector b_1 is the outward pointing normal of C in the $\{x_1, x_2\}$ plane.

B. Model-based Segmentation of the Mandible

In our segmentation method the surface model is deformed to match the edges of the reconstructed mandible. Three important features of a dental CT reconstruction that a successful model-based segmentation method has to address are:

1. The global size and shape of the mandibular arc can differ greatly from the local size and shape of the orthogonal cross-section of the mandible,
2. The reconstruction can exhibit significant and widespread artifacts due to amalgam fillings and implant screws made of metal, and
3. The appearance of the dental region can vary greatly due to, e.g., present or missing teeth and/or implant materials.

We have tackled these issues by developing an affine transformation with decoupled global and local DOFs, an adaptive outlier exclusion scheme, and by identifying the dental region $\mathcal{S}_{\text{dent}}$ of \mathcal{S} , and treating it distinctly.

We perform a series of transformations following the guidelines of a global-to-local approach. The model is first transformed affinely to capture the rough appearance and location of the reconstructed mandible, and then elastically to refine the segmentation result.

The affine stage combines a global transformation of C to describe the overall size and shape of the dental arc, and a local transformation of \mathcal{S} in the normal planes of C to capture the attributes of the mandible cross-section. The global transformation of C consists of 3D rotation and translation, and x_1x_2 scaling. The local transformation of \mathcal{S} consists of rotation about the tangential vectors of C and 2D scaling in the normal planes spanned by x_3 and b_j (see Fig. 1). This enables adjusting the attitude and shape of the surface cross-section. The combined transformation effectively captures the overall appearance of the mandible using a relatively small number of DOFs.

After the affine stage, the surface \mathcal{S} is deformed elastically by using its nodes as control points. They are displaced one-by-one in the direction of their surface normal vectors and the vector displacement is distributed smoothly to the adjacent nodes [11].

To find the optimal transformation parameters, a scalar energy measure computed from \mathcal{S} is minimized. The measure consists of the surface outward normal gradient $\mathbf{v} \cdot \nabla I$ and value I of the gray scale intensity at the surface nodes [11]. An outlier exclusion scheme is incorporated to make the computation more tolerant to outlier artifacts. It is based on approximating the values of $\mathbf{v} \cdot \nabla I$ and I during the previous energy-minimizing iteration by normal distributions and retaining the 3σ confidence interval in the computation of the energy measure. This makes the outliers effectively invisible to the model, thus making the segmentation tolerant of bright metal artifacts.

Special attention is paid to the nodes $\mathcal{S}_{\text{dent}}$ of the model (see Fig. 1), because there $\mathbf{v} \cdot \nabla I$ is likely to vanish due to teeth protruding the mandible, or be positive due to extramandibular metal artifacts. To accommodate this, the negative values of $\mathbf{v} \cdot \nabla I$ in $\mathcal{S}_{\text{dent}}$ are multiplied by 2 and the positive values by $-1/2$. The rationale is that the generally small regions, where the value is negative are emphasized to drive the model more widely towards the desired location, and the positive values are reversed in sign but attenuated to

prevent them from disturbing the convergence. This modification, albeit somewhat heuristic, clearly improved the segmentation result.

C. Synthetization of Radiographs

We defined a nonlinear transformation of $\mathcal{S} = s_i = \{s_{i1}, s_{i2}, s_{i3}\}$ in the model space M of Fig. 1 to $s_i'' = \{s_{i1}'', s_{i2}'', s_{i3}''\}$ in a panoramic space P specified here. The objective was to rectify the mandibular surface, so that C , a parabola in M became a line in P . The model coordinate system $\{x_1, x_2, x_3, o\}$ was also used in defining P . This transformation had the desirable property of re-orienting the non-parallel normal planes $\{x_3, b_j\}$ of C in M to parallel P .

The x_1 coordinate s_{i1}'' in P was defined as an Euclidean distance from o along C to the plane $\{x_3, b_j\}$ intersecting $\{s_{i1}, s_{i2}\}$ in M . The x_2 coordinate s_{i2}'' was the shortest Euclidean distance, from $\{s_{i1}, s_{i2}\}$ to C in along the $\{x_1, x_2\}$ plane intersecting $\{s_{i1}, s_{i2}\}$ in M (see Fig. 2). No transformation was done along x_3 axis, thus $s_{i3}'' = s_{i3}'$.

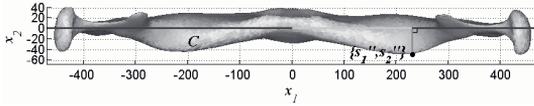


Fig. 2. The model surface (Fig. 1) was transformed from the model space M to the panoramic space P . As a result C was rectified from a parabola in the model space to a line on the x_1 axis in the P . Consequently all planes orthogonal to C in M become parallel in P .

For the purposes of synthetizing the panoramic image, we defined an isotropic grid $g_i'' = \{g_{i1}'', g_{i2}'', g_{i3}''\}$ which was aligned according to $\{x_1, x_2, x_3, o\}$. The grid covered a rectangular volume around the segmented surface in P and was used to mark the voxel locations of a panoramic volume used in the synthetization phase.

The depth and direction the teeth roots are of critical interest in dental implant planning. Due to the anatomy of the mandible, the roots are not parallel. Thus it is impossible to have an optimal panoramic view showing the whole set of tooth with roots using fixed imaging or synthetizing geometry. For this reason, we added an extra degree of freedom by rotating every plane g_k'' of g_i'' orthogonal to C according to the angle α_k of estimated tooth orientation (see Fig. 3).

We estimated the tooth orientation by studying the local shape of the segmented surface. This was done by dissecting the surface to k 2D contours matching the position and orientation of g_k'' . Several features, such as the principal direction of the contour and surface tangents of the back and front sides of the contour were extracted. These features were analyzed with linear regression with the true tooth orientation as the dependent variable. The back tangent γ_k and the distance d from o turned out to be significant ($p < 0.05$) which lead to model

$$\alpha_k = \beta_{k1}d + \beta_{k2}\gamma_k. \quad (1)$$

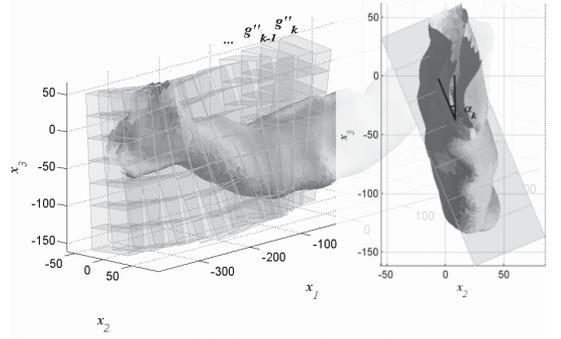


Fig. 3. An extra degree of freedom was added to align the columns of the panoramic image by rotating g_k'' according to tooth orientation α_k . The figure shows the grid g_i'' in artificially sparse resolution for the purposes of visualization.

Several more complex and nonlinear models were tried but they failed give notably better results. Parameters β_{k1} and β_{k2} were solved using least squares estimation and the validity with independent data was tested with cross-validation.

After the estimation g_k'' were rotated according to α_k and transformed first to the model space g_i and then to the image space g_i' (see Fig. 4) on the original image volume.

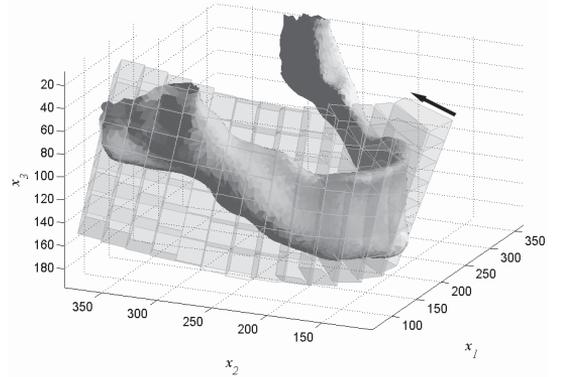


Fig. 4. The grid in Fig. 3 was transformed to the original image space. The panoramic radiograph was synthetized by summing voxels on user-selectable layer thickness and orientation in g_i' . The voxels were summed the orthogonal direction (indicated by the black arrow) within the rectangles visible in the image. A rectangle in the images corresponds to a pixel in the 2D radiograph.

The gray values of voxels of the panoramic volume were given by tricubic interpolation around points g_i' . The final panoramic image was created by adding together the layers of voxels in orthogonal (forward to back) direction of user selectable depth and thickness.

III. RESULTS

The automatically generated segmentation results were validated by comparing them to manual segmentations of the bone surface and the centerline. The average symmetric Euclidean distance between the surfaces, and correspondingly between the centerlines, was computed

according to [12]. In addition, the standard Dice similarity coefficient, measuring the volumetric overlap of the segmented objects, was defined. The average value and the standard deviation of the distance between the mandible surfaces was 0.57 ± 0.16 mm and the distance between the centerlines was 0.75 ± 0.24 mm. The average and standard deviation of the Dice coefficient was 0.91 ± 0.02 .

The correlation coefficient ρ of the teeth root angle estimation model (1) was 0.85 with std. error of 7.79 degrees using all data $N=112$. The cross-validation of cases gave the average error of 0.65 ± 8.0 degrees. The quality of the synthesized radiographs was validated visually. Fig. 5 represents a typical result of synthetization.

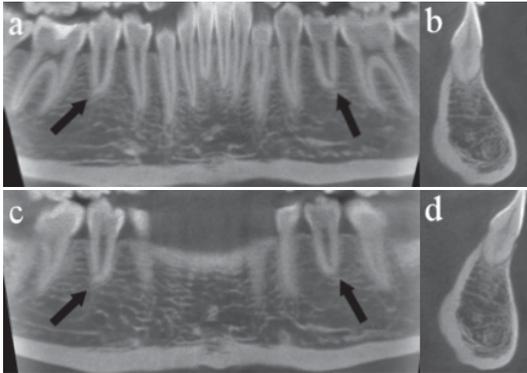


Fig. 5. Panoramic images were synthesized with (a) and without (c) the estimation of tooth orientation. A slice of thickness 0.9 mm showing premolars (marked with black arrows) was selected. This very thin slice (a) shows intersections of all teeth in a very thin slice, while in (c) incisor (front teeth) and most of the molars (back teeth) are missing. Axial slices (b,c) show the difference in orientation.

IV. DISCUSSION AND CONCLUSION

The segmentation results were very good considering the amount of artifacts and variability exhibited by the dataset. This model-based approach was developed with the application of computing simulated radiographs in mind, but the results turned out to be accurate enough for stand-alone segmentation applications.

The accuracy of the estimation of tooth orientation was sufficient for creating the panoramic views showing the whole set of teeth using a very thin slice. More accurate estimation would likely require the segmentation of the teeth individually. However, often the teeth may have been removed prior to imaging. In these cases the use of an estimation method based on bone shape based is the only possibility.

One evident topic for further development is to expand these methods for maxillary segmentation and image synthetization. This would require a model for the structures of the area, namely sinuses and the cleft palate. The principles of this method would likely work for maxillary or combined mandibular/maxillary imaging.

We used the result of segmentation in synthetizing orthopantomographs using CBCT volumes. With the method it was possible to have the useful overall view of panoramic images without the usual problems of superposition and loss of depth information. These problems were eliminated by creating a panoramic volume aligned according to individual anatomy. The panoramic radiograph was created by adding layers of voxels in the volume of used-selectable depth and thickness. Even though this navigation step still required manual work, now it could be done off-line, fast and conveniently.

As a result, it was possible to point the locations of dental structures of interest with great accuracy. This property is paramount when extracting critical measures for planning dental and maxillofacial surgeries.

REFERENCES

- [1] Farman AG, Levato CM, and Scarfe WC. "A primer on Cone Beam CT," in *Inside Dentistry*, vol. 1, pp. 90-93, 2007.
- [2] Kobayashi, K., Shimoda, S., Nakagawa, Y., Yamamoto, A.: "Accuracy in Measurement of Distance Using Limited Cone-Beam Computerized Tomography," in *Int. J. Oral Maxillofac Implants*. vol. 19, pp. 228—231, 2004.
- [3] GE Healthcare, <http://www.gehealthcare.com>
- [4] Siemens Medical Solutions, <http://www.medical.siemens.com>
- [5] Philips Medical Systems, <http://www.medical.philips.com>
- [6] Materialise, <http://www.materialise.com>
- [7] Cucciara, R., Lamma E., Tommaso, S.: "An image analysis for automatically re-orienting CT images for dental implants," in *Computerized Medical Imaging and Graphics*. vol. 28, pp. 185—201, 2004.
- [8] Rueda, S., Gil, J. A., Pichery, R. and Alc niz, M.: "Automatic segmentation of Jaw Tissues in CT Using Active Appearance Models and Semi-Automatic Landmarking," in: Larsen, R., Nielsen, M., Sparring, J. (Eds.): *MICCAI 2006*, LNCS vol. 4190, pp. 167–174. Springer-Verlag, Heidelberg 2006
- [9] Lamecker, H., Zachow, S., Wittmers, A., Weber, B., Hege, H-C., Isoholz, B., Stiller, M.: "Automatic segmentation of Mandibles on low-dose CT-data," in *Int J CARS* 1, 393–392, 2006
- [10] Humphries, S., Christensen, A., Bradrick, J.: "Cone beam versus conventional computed tomography: comparative analysis of image data and segmented surface models," in *Int. J. CARS* 1, pp. 389–402 (2006)
- [11] M. Lilja, V. Vuorio, K. Antila, H. Set l , J. J mstedt, M. Pollari: "Automatic segmentation of the mandible from limited-angle dental x-ray tomography reconstructions," in *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. USA, 12–15 April pp. 1933 – 1936, 2007
- [12] Niessen, W. J., Bouma, C. J., Vincken, K. L., Viergever, M. A.: "Error Metrics for Quantitative Evaluation of Medical Image Segmentation," in *Proceedings of the Theoretical Foundations of Computer Vision, TFCV on Performance Characterization in Computer Vision*, pp. 275–284. Kluwer, B.V. 2000.

PUBLICATION IV

**Segmentation of facial bone surfaces by patch growing from cone beam CT
volumes**

Kari Antila, Mikko Lilja and Martti Kalke

Dentomaxillofacial Radiology, 45(8), p. 20150435
<https://doi.org/10.1259/dmfr.20150435>

Publication reprinted with the permission of the copyright holders.

RESEARCH ARTICLE

Segmentation of facial bone surfaces by patch growing from cone beam CT volumes

¹Kari Antila, ²Mikko Lilja and ³Martti Kalke

¹VTT Technical Research Centre of Finland, Espoo, Finland; ²Department of Biomedical Engineering and Computational Science, School of Science, Aalto University, Espoo, Finland; ³KaVo Kerr Group, Tuusula, Finland

Objectives: The motivation behind this work was to design an automatic algorithm capable of segmenting the exterior of the dental and facial bones including the mandible, teeth, maxilla and zygomatic bone with an open surface (a surface with a boundary) from CBCT images for the anatomy-based reconstruction of radiographs. Such an algorithm would provide speed, consistency and improved image quality for clinical workflows, for example, in planning of implants.

Methods: We used CBCT images from two studies: first to develop ($n = 19$) and then to test ($n = 30$) a segmentation pipeline. The pipeline operates by parameterizing the topology and shape of the target, searching for potential points on the facial bone–soft tissue edge, reconstructing a triangular mesh by growing patches on from the edge points with good contrast and regularizing the result with a surface polynomial. This process is repeated for convergence.

Results: The output of the algorithm was benchmarked against a hand-drawn reference and reached a 0.50 ± 1.0 -mm average and 1.1-mm root mean squares error in Euclidean distance from the reference to our automatically segmented surface. These results were achieved with images affected by inhomogeneity, noise and metal artefacts that are typical for dental CBCT.

Conclusions: Previously, this level of accuracy and precision in dental CBCT has been reported in segmenting only the mandible, a much easier target. The segmentation results were consistent throughout the data set and the pipeline was found fast enough (<1-min average computation time) to be considered for clinical use.

Dentomaxillofacial Radiology (2016) **45**, 20150435. doi: 10.1259/dmfr.20150435

Cite this article as: Antila K, Lilja M, Kalke M. Segmentation of facial bone surfaces by patch growing from cone beam CT volumes. *Dentomaxillofac Radiol* 2016; **45**: 20150435.

Keywords: computer-assisted image analysis; CBCT; dental implantation

Introduction

Precise planning of dental implants and dental and maxillofacial surgeries require volumetric (three-dimensional) images from the area to be operated. CBCT has been developed as a relatively low-cost and low-dose alternative to conventional CT to meet these needs.¹ During the past decade, CBCT has become an established radiologic technique in dental imaging.

The full advantage of volumetric images will be obtained by visualizations of the target and the neighbouring structures, such as a tooth to be replaced with

an implant and the inferior mandible canal. These include (two-dimensional) cross-sectional slices of the dental arch, panoramic radiographs of chosen geometry and depth and volumetric renderings.² When these two-dimensional images or views are reconstructed from volumetric images, their correspondence to anatomical ground truth can be retained and the reconstructed image orientation, location and geometry fitted to individual anatomy when needed. Creating these views typically require segmentation of the anatomic structures from the volumetric images. Image segmentation is a very useful and important tool in other purposes also, such as in the evaluation of tumorous bone infiltration, intrabony

Correspondence to: Mr Kari Antila. E-mail: kari.antila@vtt.fi

Received 29 December 2015; revised 8 July 2016; accepted 13 July 2016

pathologies, fracture diagnostics and orthognathic monomaxillary and bimaxillary surgery planning.^{3–5}

When image segmentation is performed manually, the results are hard to reproduce. Manual drawing on potentially hundreds of slices of high-resolution image volumes takes considerable amount of time. Assuming that manual drawing of the facial contour takes 10 s per slice, drawing every slice of a single image consisting of 300 slices would take 25 min. Thus, automatic segmentation methods would be much preferred. Unfortunately, CBCT or CT volumes often present strong metal and other artefacts.⁶ The lower dose of CBCT yields a lower signal-to-noise ratio, worse contrast and higher intensity inhomogeneity than conventional CT. The highly varying anatomy of the mandible and maxilla, especially around the teeth, poses another major challenge.

Lamecker *et al*⁷ were one of the first to report results on segmenting the mandible from CBCT volume. Their strategy was to use a statistical shape model, which was deformed around the mandible.⁸ Rueda *et al*⁹ developed a method for segmenting the cortical bone and other targets from the cross-sectional slices of the mandible. They exploited both the shape and texture of the hand-segmented structures to train an active appearance model of the structures of interest.¹⁰ Kainmueller *et al*¹¹ extended the work by Lamecker *et al*⁷ by improving the segmentation method and extending the application to tracking the mandibular canal. Recently, Wang *et al*¹² have presented a volumetric segmentation method for both the mandible and maxilla. Their method registers a number of prior models (atlases) to the target and generates patient-specific prior model constructed of patches that are selected from the registered images.

Past works on segmenting dental CBCT volumes concentrate mostly on segmenting the mandible and its internal structures. Shape and appearance models used by published algorithms do not include the maxilla or teeth, which are the structures most prone to exhibiting metal artefacts, and usually also have the highest variation of shape and consequently represent areas where segmentation algorithms are most likely to fail. In order to develop an algorithm capable of segmenting the exterior of these structures, several established segmentation paradigms were considered. Registering the volume according to a mean intensity model was deemed unsuccessful owing to the high variation in anatomy and variation in pose and the varying state of the mandible (mouth) being closed or open to some extent.¹³ Atlas-based methods have been applied successfully in various applications such as brain image analysis and now also dental CBCT.^{12,14,15} However, generation of the shape and intensity models as required by atlases would have been challenging owing to the highly varying anatomy, pose and scanned area of our population. The usually very high computational cost and long execution times associated with atlas-based segmentation tools were unacceptable for our clinical tool. Active appearance models, statistical shape models or other deformable

models have been found to be successful in segmenting the mandible.^{2,7,11} However, these models require a rather accurate initialization. In the typical case where the exact location, orientation or delineation of the target relative to the volume is not precisely known, the initialization of these models would sometimes require laborious manual interaction.

The aim of our study was to develop an automatic and data-driven segmentation algorithm that requires only very general and non-specific knowledge of the target. The algorithm was specified to segment the exterior of all visible bones in a CBCT volume with a single open surface in three dimensions. The algorithm was required to be accurate, precise and fast enough for clinical use and to tolerate the challenging characteristics of dental CBCT and possible gaps and holes around the facial skeleton. The intended use for the algorithm was to aid in creating visualizations and thus, good, consistent and continuous overall fit to the target was preferred over highly detailed result over small and local shapes.

Methods and materials

Data

The data set consisted of 49 isotropic CBCT volumes collected by scanning human subjects during two separate studies S1 ($n = 19$) and S2 ($n = 30$) on separate occasions. The subjects were scanned using manufacturer Soredex Scanora model (KaVo Kerr Group, Tuusula, Finland) prototypes during their development. S1 was scanned using an early version of the system and prior to S2, the system was upgraded. The volumes were reconstructed in cylinder-shaped fields of view (FOVs) in three different sizes (diameter \times height: 60×60 , 100×75 and 145×75 mm). The 145×75 -mm FOVs were available only in S2, after the upgrade. The volumes were isotropic sized from $300 \times 300 \times 300$ to $580 \times 580 \times 300$ voxels and from 0.13 to 0.35 mm per voxel. All subjects were patients scanned by an authorized healthcare provider according to rules of ethics. No normal controls were included because of the use of ionizing X-rays in CBCT. Subject diagnoses or any other personal information was not disclosed with the images. Most volumes had at least one of the following types of artefacts: metal, movement, inhomogeneity or noise, some severe (Figure 1). Roughly, 36.7% of subjects had their mouth open to some extent during scanning (Table 1). Most of the subjects were presumably candidates for implant surgery and thus were missing some, even most, of the teeth. Positioning information, such as the anatomic region of interest (mandible, maxilla or sinuses) or the placement of the FOV (left, centre or right), was not included with the images.

Overview of the segmentation algorithm

The basic principles of the developed segmentation algorithm are rather straightforward: finding a sufficiently

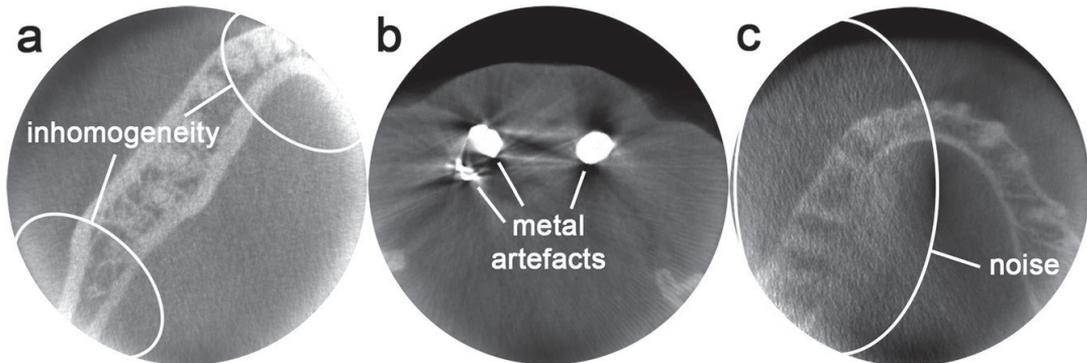


Figure 1 Artefacts typical for dental CBCT: CBCT volumes are acquired with a smaller X-ray exposure and thus have a worse signal-to-noise ratio when compared with conventional CT. The algorithm was designed to segment volumes with inhomogeneity (a), metal artefacts (b), movement and noise (c) typical for dental CBCT.

large number of points with good coverage on the exterior bone–soft tissue edge of the facial skeleton and reconstructing a surface mesh on them. The algorithm was implemented in a pipeline of sequentially performed steps (Figure 2).

The pipeline begins with filtering to suppress noise and to estimate a number of parameters to be used later in the pipeline. The main novelty and the key elements of the algorithm are in four steps following filtering. First, the coarse shape and location of the target is parameterized. This parameterization also defines a topology according to which a set of equally spaced line profile segments presumably intersecting the target edge is aligned. Second, potential edge points are searched along these segments and the best candidates are selected. Third, a mesh of triangles connecting the neighbouring edge point candidates on the segments is reconstructed and the best triangles are selected for the final mesh configuration. Fourth, a surface polynomial is fitted to the mesh to fill holes and gaps and to regularize and smooth the result to a chosen degree. These four steps are iterated and once a good fit to the target is reached, the iteration loop is terminated and the resulting surface mesh is the output. The following sections discuss these steps in detail.

Filtering

The input volume is first filtered with a three-dimensional, $\sigma = 1.5$ -mm Gaussian kernel. A kernel of

this size is large enough to smooth the noise typical of our data set but not small enough to blur the target edge too much or lose other relevant detail. Kernel size may have to be determined for different modalities, image resolutions, quality and targets. The volume is filtered again with a two-dimensional Sobel-type kernel to estimate the magnitude and orientation of the edge gradient for every voxel in the volume. The kernel is applied to all slices of the image volume in axial orientation.

The approximate intensity ranges of the bone and the surrounding soft tissues need to be estimated for later use in the segmentation pipeline. In CT, a natural way of achieving this would be to use known ranges for these tissues in Hounsfield units. Unfortunately, owing to different volume reconstruction techniques used in CBCT, this is not necessarily possible and mapping the CBCT intensities to the Hounsfield units scale may not be straightforward. For these reasons, we estimate the ranges with a direct clustering-based classifier (Figure 3).¹⁶ Four predefined classes were set to approximate (1) the background, (2) soft tissue, (3) bone and (4) hard objects such as the tooth enamel and metal. The means and standard deviations of the intensities of the voxels labelled in the soft tissue and bone are retained.

Parameterization

The coarse shape, location and orientation of the target are captured by an arc length-type parameterization.

Table 1 General characteristics of the data set

Set	Mouth open	Maxilla visible	FOV centre	FOV left	FOV right	Noise	Metal
S1	15.8%	89.5%	89.5%	0.0%	10.5%	42.1%	63.2%
S2	50.0%	76.7%	76.7%	20.0%	3.3%	0.0%	70.0%
All	36.7%	81.6%	81.6%	12.24%	16.3%	16.3%	67.4%

FOV, field of view; S1, Study 1; S2, Study 2.

The table summarizes the relative number of images in S1, S2 and all where the subjects had the mouth open, the maxilla was at least partly visible, how the FOV was placed (centre, left or right) and whether the image exhibited metal or noise-type artefacts. This information was compiled by inspecting the data set visually.

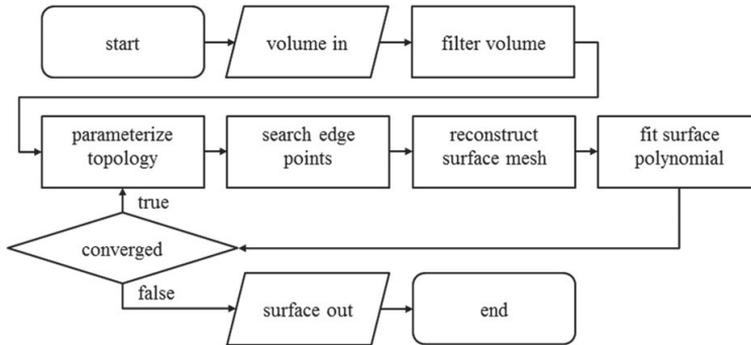


Figure 2 Overview of the automatic segmentation CBCT pipeline: the input volume is read and filtered to suppress noise, extract image gradients and estimate parameters of the volume content. The four steps: parameterizing the surface topology, searching for potential target edge points, reconstructing surface edge points and fitting a surface polynomial to the reconstructed mesh are iterated for best results. Once converged, the resulting triangular surface mesh is provided as output.

Arc length parameterization works well on the characteristic shape of the dental arch, although other approaches such as spherical parameterization could be used.¹⁷ The parameterization is obtained by mapping the points $x = \{i, j, k\}$ of the isotropic Cartesian index coordinate system of the image voxel grid to a parameterized system $x' = \{i', j', k'\}$, where the coordinates are the signed distance from the arc i' from the apex and the signed distance j' from the nearest point on the arc. Coordinate k is the axial slice index and remains unchanged by the mapping, *i.e.* $k' = k$. Only one arc is defined for the whole image volume. This mapping is illustrated in Figure 4.

The arc is defined by a third-degree polynomial. Its coefficients are solved by fitting the polynomial to edge points of the whole surface projected on the $\{i, j\}$ (axial) plane. At the start when no edge points exist yet, the

polynomial is simply a horizontal line from left to right, cutting the $\{i, j\}$ plane in half. As the pipeline runs further, the arc converges towards the dental arch and facial bones. This parameterization defines a plane according to which property segmentation surface will be defined.

The parameterization plane defined by the arc will not have enough degrees of freedom to fit to the target exactly and it is not supposed to do so. The parameterization is meant to only capture the global curvature of the target and to define a rough topology and space where potential target edge points are searched.

The edge points are searched on a set of line profiles along orientation j' in the parameterized space. A number of line profiles are placed on a grid in the $\{i', j'\}$ plane on a defined spacing. Uniform spacing of 1.25 mm was used for the best combination between accuracy and computation time. The grid also defines

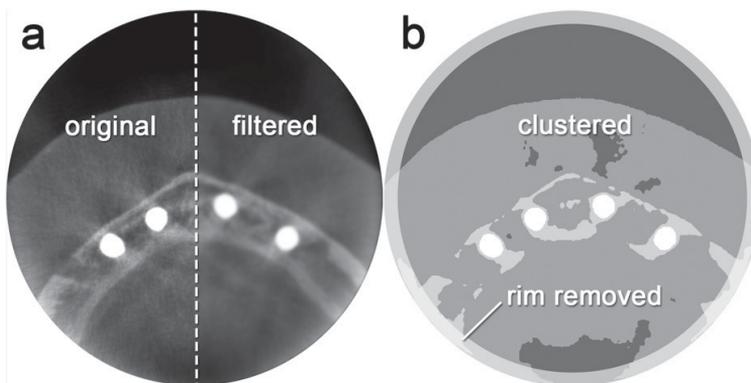


Figure 3 A CBCT image is filtered and the image voxels clustered according to thresholds defined by a direct-clustering method to approximate intensity ranges of the tissue classes before segmentation. Noise and small artefacts are smoothed with Gaussian filtering (a). The filtered images are clustered into classes of (1) background (dark grey areas), (2) soft tissue (grey areas), (3) bone (light grey areas) and (4) hard objects such as enamel and metal (white areas) (b). A small portion of the outside rim of the field of view (FOV) is removed from the parameter estimation owing to regular inhomogeneity in the area (b) as if the bone would spread along the borders of the FOV at 4 and 8 o'clock orientations.

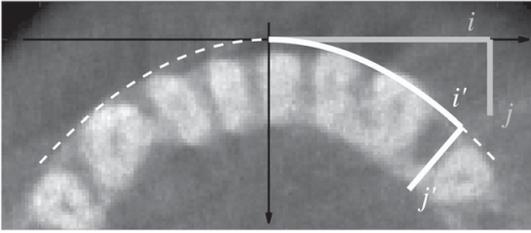


Figure 4 Image grid coordinate system and the parameterized system: the image coordinate system $\{i,j,k\}$ is an isotropic, Cartesian grid where coordinates $\{i,j\}$ lie on the axial plane. Parameterization bends the original coordinate system (grey areas) along a polynomial (white areas) fitted on the outer surface of the dental arc. The resulting coordinates of the parameterized space are signed arc length from the apex i' and signed distance to the nearest point on the arch j' . The mapping is two dimensional and thus $k' = k$.

a topology in which the actual surface is reconstructed during the later stages of the pipeline. Use of line profiles has the benefit of limiting the space and the number of voxels for the search to achieve faster computation.

Edge point search

Potential edge points are searched by estimating an energy as follows:

$$E(x') = a \left| \frac{g(x')}{\left(\frac{v_{\text{diff}}}{r}\right)} - 1 \right| + b \left| \frac{v(x') - v_n}{v_{\text{diff}}} \right| + c \left| \frac{o(x') - o_n}{\frac{\pi}{2}} \right|, \quad (1)$$

for every voxel x' crossing a line segment at $n = \{i', k'\}$, where $g(x')$, $v(x')$ and $o(x')$ are the gradient magnitude, image intensity and gradient orientation, respectively. $E(x')$ has similarities to energy functions used in other segmentation methods but was defined specifically for this problem.^{2,7,9,11} $g(x')$, $v(x')$ and $o(x')$ are estimated

at the filtering step of the pipeline and remain constant. v_{diff} is the difference between the average intensities of voxels labelled to the soft tissue and bone and r is the radius of the Sobel kernel used in estimating edge gradients. v_n and o_n are the estimated intensity and target orientation of the line profile segment at n . v_n and o_n are updated at every iteration based on the intensities and orientations of the edge point at n .

a , b and c are weighing constants determined with machine learning or, in this work case, determined simply by exhaustive search by using the training set. $E(x')$ was designed as a minimum energy function with the three components normalized to the scale of $[0,1]$ with the value of 0 indicating the best possible properties.

Once the energy values have been computed for every voxel crossing a line profile, a number of q voxels at local minima per every profile are selected. In this study, we used $q = 3$, as in the vast majority of the cases, the correct edge was found among the three best candidates. The search of local minima along the line segments is illustrated in Figures 5 and 6.

Surface mesh reconstruction by patch growing

Besides the actual segmentation, the aim of reconstructing the target surface is to determine which of the edge point candidates lay on the same edge and thus are part of the same structure or object and further which surface represents the desired facial bone–soft tissue edge. The surface is constructed as a triangular mesh whose topology, *i.e.* the node to vertex configuration, is fixed at the parameterization step and remains constant. The goodness of fit of the mesh to the edge will be estimated locally by computing energies for the edge point candidate–triangle combinations using Function (1). The energies are estimated for every image voxel crossing the triangle and the median is taken by giving

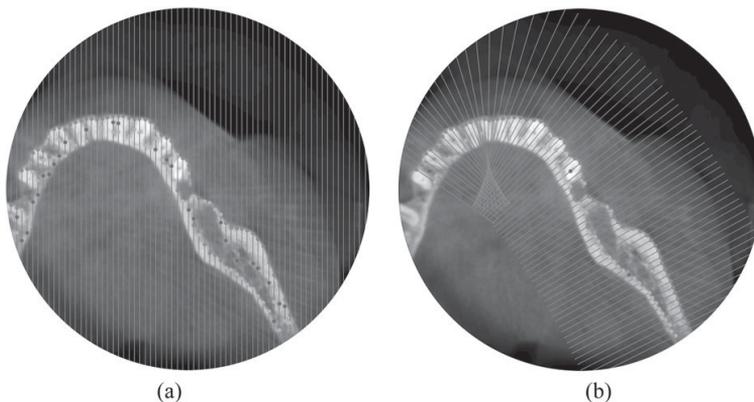


Figure 5 Edge point search along the line profiles: potential edge points are searched along the line profiles and a number of q (here, $q = 3$) best candidates are selected for further consideration. The best candidates are shown in red, second best in green and third best in blue. Image (a) shows the line profiles and point candidates after the first iteration when the parameterization and thus the last component of the energy Function (1) is not yet used. Image (b) shows a slice with line profile configuration bent along the parameterization during later iterations where segment lengths are trimmed for shorter energy profiles and thus less voxels for evaluation. For colour image see online.

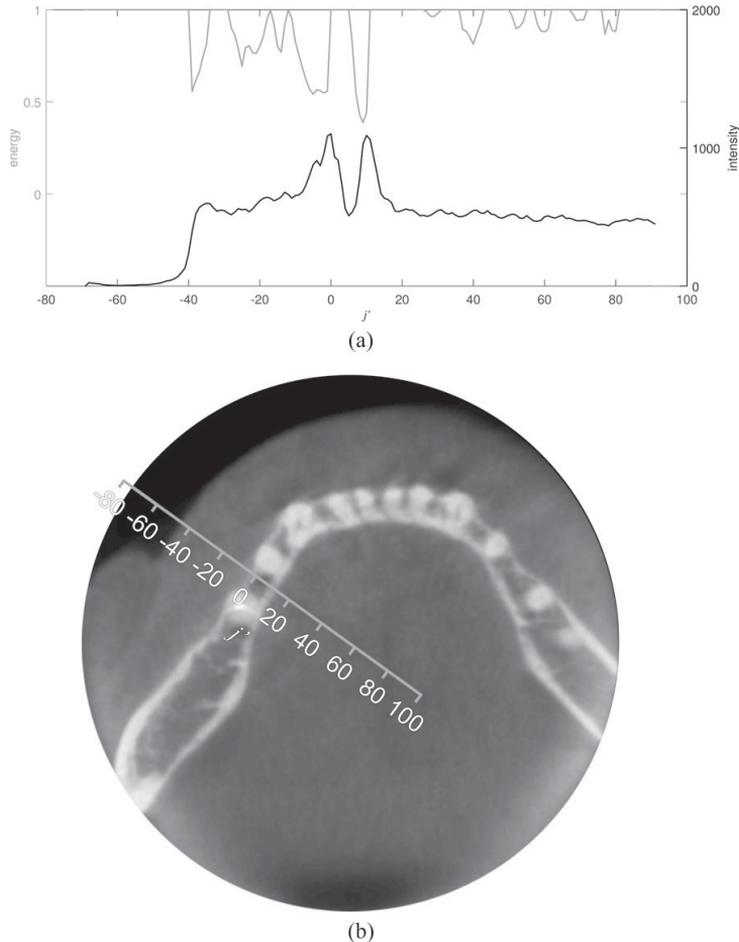


Figure 6 Energy and intensity: corresponding energy (grey line) profiles as computed with Equation (1), with the corresponding intensity (black line) (a). The profiles show that the energy minima correspond to the highest gradients in the intensity that match the edges of the outer and inner surfaces of the mandibular bone. The location and orientation of the intensity profile in (a) is shown on an axial image slice in (b).

one energy value per triangle. The triangles with the lowest energy will be selected for the mesh if a maximum energy threshold of 0.9 is met. In case of exceeding the energy threshold, the triangle is discarded, leaving a hole in the mesh.

To find a global minimum, all triangle combinations by neighbouring point candidates should be tested. This, however, would lead to an unnecessarily high computational cost, since the energy values for triangles of q^3 edge points would need to be estimated for every triplet of line segments. Instead of going through all combinations, a number of seed points are selected and fixed. All $(q^2 \times 4)$ triangle combinations around the seed and the adjacent edge point candidates are evaluated and the minimum energy triangle is taken. This single triangle acts as the start of a surface patch

(Figure 7) that is further grown by fitting the triangles on the patch boundary to the edge point candidates on the neighbouring line segment and adding the minimum energy triangle to the patch if it falls below the maximum energy threshold. Since all but the single tested edge point candidate (node) is fixed to the patch already, only the maximum of q combinations per triangle need to be tested. Growing patches in this manner does not guarantee a global minimum for the whole mesh but gives a very good chance of finding the local, consistent edges and requires only linear time to compute.

The criteria for the selection of seed points are critical for providing an accurate starting point for the patch-growing phase and thus successful surface mesh reconstruction. We chose to select the points that have the

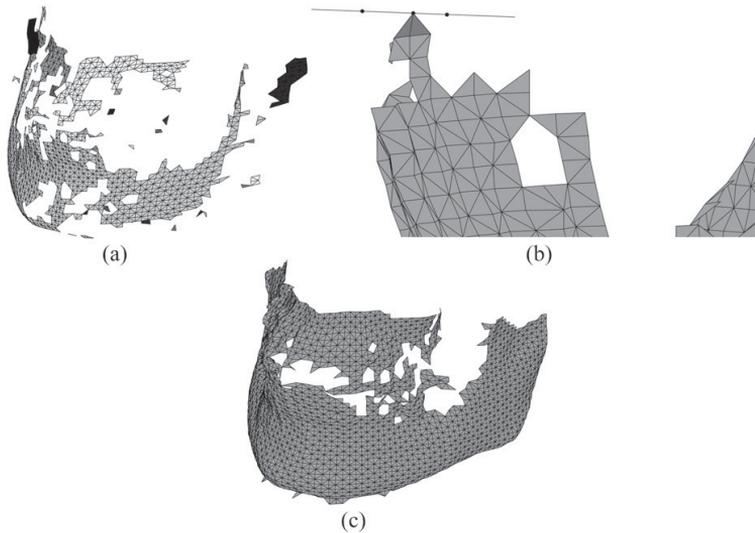


Figure 7 Growing surface patches to a mesh: a number of seed points potentially on the bone–soft tissue edge are selected and a number of patches (separate patches in different shades of grey) are formed by joining neighbouring seed points by triangles (a). The patches shown are grown by adding potential edge points on the analyzed line segment (b). Image (c) shows the resulting surface mesh of the patch-growing phase.

lowest energy and lie foremost on their segments. These are most likely correct for segmenting the outer surface of the facial skeleton, since the only edges in dental CBCT outside the skeleton are the soft tissue–air boundary or artefacts. In another application, some other criteria could be used or the operator could be asked to select a number of points depending on the size and contrast of the target to act as seeds.

The patch growing will be run for every surface patch independently. This will result in a number of patches, some of which might overlap. In other words, two edge point candidates from the line segment may belong to different surfaces. This will lead to a problem where the right patches to represent the target surface need to be selected. In this application, the two largest non-overlapping patches were taken. It was simply presumed that the two largest patches represent the maxilla and mandible. Also, additional rules such as the minimum patch size and maximum average triangle energy threshold were used to discard patch edges likely other than facial bones (sinus cavities *etc.*) from consideration.

Fitting of surface polynomial

The surface reconstruction will result in a mesh with holes and disjoint patches. To bridge these holes and gaps and to obtain smooth and consistent surfaces, the mesh needs to be interpolated. This is performed by fitting a thin plate smoothing spline polynomial $f(n)$ to the edge point candidates that remain on the mesh patches after the growing phase.¹⁸ The smoothing is performed in the parameterized space. The coefficients of f are estimated by minimizing the sum as follows:

$$S = p \times \varepsilon(f) + (1 - p) \times R(f), \quad (2)$$

where

$$\varepsilon(f) = \sum |j' - f(i, k')|^2 \quad (3)$$

is the error measure between the mesh node coordinate j' and the value of the function f at coordinate $n = \{i, k'\}$ and $R(f)$ is a roughness measure penalizing the bending of the surface too sharply. A smoothing parameter $p \in [0, 1]$ acts as the weighing term between the two components. $R(f)$ can be chosen according to the application. We used the integral of the second derivatives of the node coordinates as implemented in MATLAB curve fitting toolbox function *tpaps* (MathWorks, Natick, MA).

The polynomial is fitted to the edge points in the parameterized space, since this removes the global curvature of the target and thus enables an effective representation of the surface. Figure 8 shows a typical result of a surface fitting to a mesh just after reconstruction.

Convergence

Once the surface polynomial has been fitted, its properties will be evaluated. First, it needs to be determined whether to continue or terminate the iteration loop (Figure 2). During the early development of the algorithm, the decision was based on the estimation of the movement of the surface mesh node coordinate j' from the current and previous iterations. If 95% of the nodes deviated less than a chosen threshold from the previous iteration, the loop was terminated. Since most of the progress is typically gained during the early iterations,

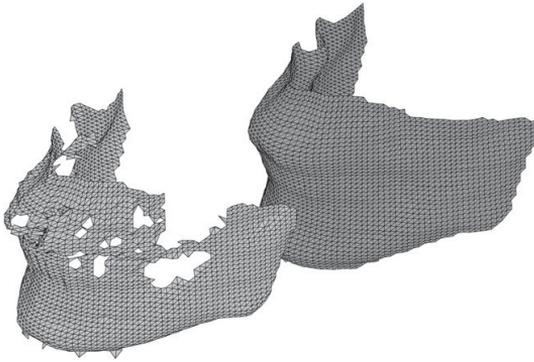


Figure 8 Surface regularization with a thin plate polynomial: a thin plate spline surface is fitted to the raw surface mesh (light grey) which typically contains holes, sharp peaks etc. The spline surface (dark grey) will fill the gaps and smooth the result to a chosen degree.

we decided to run a fixed number of three iterations instead. Iterations beyond that gave very little improvement for the cost of the extra computation time used. If the loop is continued, the parameterization polynomial will be refitted to the current edge point coordinates and the terms v_n and o_n in Equation (1) are re-estimated.

Assessment

The results were validated against a reference surface drawn manually by an expert (a medical physicist with 10 years' experience). The facial bones were drawn by placing markers on the axial slices of 2-mm spacing, covering all bone–soft tissue contours visible to the eye. The segmentation error was computed as the closest distance from the markers to the segmented surface. The distance was measured from a marker to the plane spanned by a mesh triangle along the normal direction of the plane. All marker–mesh triangle combinations were tested and the shortest distance was taken for every marker. In addition, the marker–triangle correspondence was tested by scaling the mesh (inflation or deflation around the centre of the mesh) so that it would cross the tested marker. Only those marker-to-surface measurements where the marker crosses a triangle of the scaled mesh were taken into account. This procedure also gives an estimate of how much the automatically segmented surface area covers the hand-drawn target. The average distance, standard deviation and root mean squares (RMS) of the marker–surface distances per image were estimated.

Results

Accuracy

Visual examples of the best, average and worst performance of the algorithm are shown in Figure 9.

Figure 10 shows the distributions of distances from the hand-drawn markers to the segmented surface per image.

The results per volume are shown in Tables 2–4.

The study S1 was used as a training set to find the best values for the constants a , b and c in Function (1) and S2 was used for validation. The values $a = 0.48$, $b = 0.26$ and $c = 0.26$ for Function (1) were found to give the best average accuracy. The largest weight in the energy function was thus given to edge gradient magnitude a . The relatively low values (<0.1) of p in Function (2) gave the most consistent surfaces. Thus, a rather large weight was given to the smoothing component of Function (2), smoothing the sharp curvatures.

The average computing time using an iMac OS X 10.6 (Apple, Cupertino, CA) Intel i5 2.4-GHz (Intel, Santa Clara, CA) workstation was just below a minute per volume running a single thread. The segmentation algorithm and the validation method were mostly implemented in MATLAB 2011b with a few parts such as the computation of line–voxel and triangle–voxel intersections written in C. The single most computationally intensive step was solving the thin plate surface polynomial, which took roughly half of the computation time. About one-third of the time was spent in filtering the volumes and the rest mostly in parameterization, energy computations and mesh reconstructions.

Application

The primary motivation behind developing the algorithm was to generate anatomy-based radiographs with no or very little user interaction. For example, in implant planning, a typical problem is to determine the exact location of the mandibular canal relative to the mandibular bone or the bottom of the sinus cavity relative to the maxillary bone. Here, we present an example application where a panoramic slice and a number of cross-sectional views (Figure 11) were reconstructed based on the segmentation of the facial bones on a presumed location (right molar) of the implant.

Discussion

An algorithm capable of segmenting the exterior of the facial bone surfaces including the mandible, teeth, maxilla and zygomatic bones was developed and validated. The developed algorithm reached an accuracy of 0.5 mm (averaged over all images) from the segmented surface to the manually drawn markers. The segmented surfaces covered an average of 92% of the area (markers) of the reference surfaces. Extensive coverage is a significant indicator of the performance of the algorithm, since all facial bones in the images visible to the human eye, including the most distorted and thus difficult to segment areas, were included in the reference.

The precision of the developed algorithm was evaluated by computing the RMS distance between the segmented surface and the manually drawn markers. The

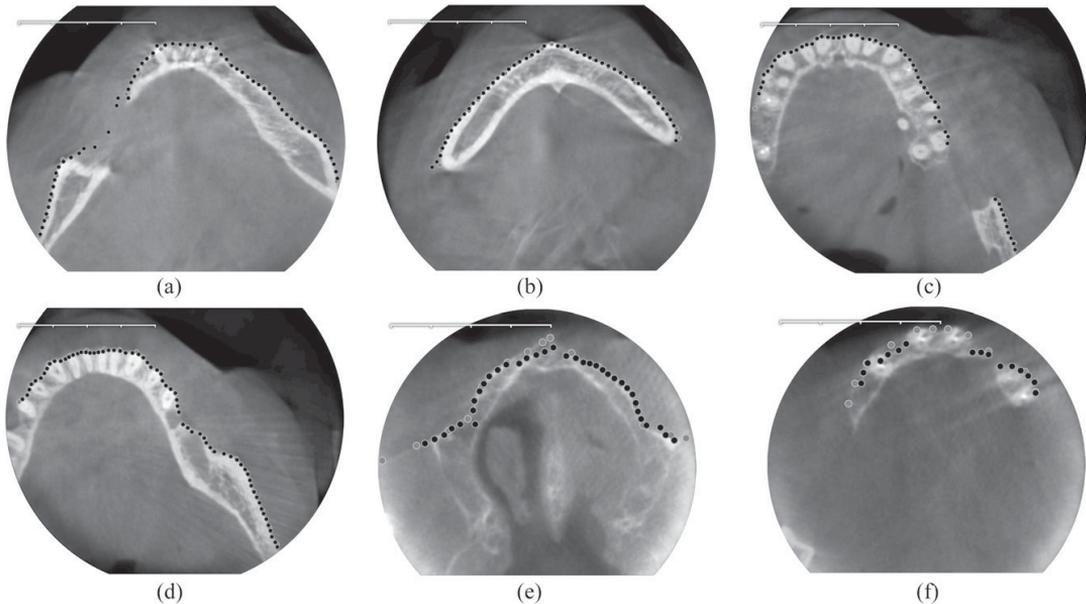


Figure 9 The best, average and worst segmentation results: slices (a) and (b) show the best result based on the shortest average distance from a hand-drawn surface. The automatically detected edge points (black circles) and manually drawn markers (grey circles) are shown on axial image slices. Slices (c) and (d) show a typical result. Some of the worst results occur when the algorithm gets attracted to false edges owing to bad contrast (e) and metal artefacts (f) or when the amount of applied smoothing is excessive to capture the sharp topological features such as the maxilla just below the zygomatic bone or a combination of these. The ruler overlay units are in centimetres.

algorithm reached an average RMS distance of 1.1 mm, the worst image giving 2.9 mm. This indicates that none of the surfaces segmented had major deviations from the reference. These results are close to the most accurate reported in dental CBCT segmentation by Kainmueller *et al.*¹¹ They achieved the average RMS distance of 0.8 mm compared with our 1.1 mm in this study. Kainmueller *et al.*'s¹¹ results were reported for the mandible, whereas ours include the exterior of the maxilla. Wang *et al.*¹² reported excellent dice ratios (0.91 for the mandible and 0.87 for the maxilla) and accuracy (average surface distance) similar to our method (0.61 vs. 0.50 mm). The average surface distance of Wang *et al.*'s¹² method was reported for the mandible. The

computational cost of Wang *et al.*'s¹² algorithm was very high (5 h) that limits its clinical application.

The developed method was found to give consistent results for the entire data set. The visual impression of the algorithm robustness in noisy images and images with weak edges was very good. The developed algorithm, like almost any other segmentation method, is still somewhat prone to false edges. In dental CBCT, there may be several false edges with an intensity profile and orientation similar to those of the correct edges owing to inhomogeneity, metal, reconstruction errors or other artefacts. The rather simple logic of our algorithm in choosing the correct edges may get distracted in areas with similar, competing edges. This is the cause for the largest

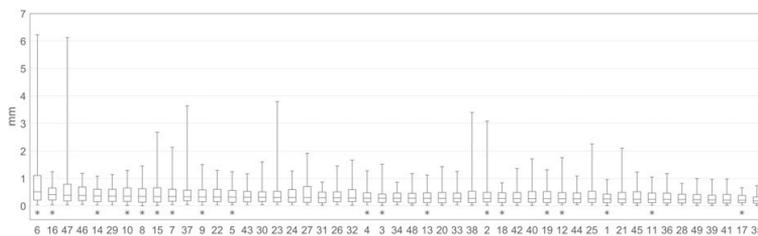


Figure 10 Distributions of surface-to-hand-drawn marker distances (errors): the distances [in millimetres (mm)] between hand-drawn markers and segmented surfaces are shown with box and whisker plots, where the midline in the box is the median, the box limits are 25 and 75% and the whisker limits are 5 and 95%. The distributions are ordered from poorest to best (left to right) as measured in the width of the 5 and 95% limits. The horizontal axis shows the image indices. Images that belong to the teaching set (Study 1) are marked with asterisks (*).

Table 2 Individual segmentation results for the Study 1 data set

Index	Mean (mm)	Std (mm)	RMS (mm)	Coverage (%)
1	0.32	0.30	0.44	92
2	0.81	2.81	2.91	76
3	0.46	0.75	0.88	88
4	0.43	0.70	0.82	85
5	0.44	0.46	0.63	86
6	1.31	2.31	2.65	81
7	0.83	2.15	2.30	85
8	0.59	1.50	1.61	94
9	0.47	0.47	0.66	90
10	0.48	0.50	0.69	97
11	0.36	0.36	0.51	90
12	0.46	0.71	0.84	94
13	0.40	0.54	0.67	91
14	0.43	0.34	0.55	89
15	0.60	0.86	1.04	90
16	0.51	0.56	0.76	98
17	0.72	2.59	2.67	78
18	0.51	2.35	2.40	92
19	0.44	0.63	0.76	89

RMS, root mean squares; Std, standard deviation.

deviations shown in Figure 10. It may not be simple to overcome this by a fully automatic algorithm in dental or other applications, but a feature where the operator manually selects one or two correct surface patches from a number of proposed ones in case of difficult images could be trivially added to the current algorithm.

The parameters of our algorithm, the weights of Equation (1), smoothing parameter in Equation (2) and spacing of the parameterization grid, were set to reach the

Table 3 Individual segmentation results for the Study 2 data set

Index	Mean (mm)	Std (mm)	RMS (mm)	Coverage (%)
20	0.43	0.61	0.74	96
21	0.51	0.87	1.01	94
22	0.45	0.48	0.66	100
23	0.87	2.29	2.44	95
24	0.46	0.63	0.78	95
25	0.56	0.93	1.08	97
26	0.49	0.80	0.94	98
27	0.62	1.13	1.29	93
28	0.31	0.28	0.41	97
29	0.56	1.69	1.78	97
30	0.46	0.54	0.71	97
31	0.34	0.29	0.45	94
32	0.49	0.67	0.83	96
33	0.42	0.57	0.71	90
34	0.34	0.26	0.43	96
35	0.25	0.26	0.36	95
36	0.36	0.39	0.53	90
37	0.77	1.22	1.44	80
38	0.62	1.10	1.26	87
39	0.33	0.43	0.55	91
40	0.47	0.58	0.75	96
41	0.35	0.55	0.65	94
42	0.41	0.52	0.66	96
43	0.40	0.36	0.54	97
44	0.36	0.35	0.50	95
45	0.53	1.44	1.53	98
46	0.48	0.38	0.61	97
47	1.11	1.89	2.19	98
48	0.39	0.51	0.64	98
49	0.33	0.40	0.52	96

RMS, root mean squares; Std, standard deviation.

minimum average distance from the reference. The best combination of these was determined by an exhaustive search over a large range using study S1. The selected combination was independently tested with S2. The fact that the algorithm performed better in S2 than in S1 is explained by the improvement in the image quality owing to equipment upgrades between S1 and S2. In fact, S1 represents early prototype data, which provided us the opportunity to test our method in a much more challenging environment when compared with today's requirements for image quality. On the other hand, many images in S2 had larger FOVs that also included the zygomatic bones that are sharper in shape than those present in the smaller FOVs of S1. Since the choice of the roughness measure and value for the smoothing parameter in Function (2) was solely based on S1, we saw somewhat poorer performance in some of the large FOVs of S2. We would have likely achieved the best results by estimating the parameters with a cross-validation scheme that uses images from both studies, but since we had already used all images in S1 at the early design stages of the algorithm, this was not possible.

The thin plate spline surface polynomial was chosen to regularize and extend our segmentation surface over areas that have holes or are severely affected by artefacts. The use of the smoothing term with the spline results in a trade-off between the average and RMS distances. In general, using less smoothing gives a better accuracy but worse precision. Aggressively smoothed results would be beneficial for applications such as panoramic and cross-sectional slice reconstruction, where sharp curves could complicate the calculation of the orientations of the cross-sections and the location of the sharp layer. On the other hand, applications such as segmentation for volume rendering could benefit from a less smooth and thus more detailed fit to the target.

The primary motivation of the work was to develop a segmentation method to be used in the automatic reconstruction of cross-sectional and panoramic views of the facial and dental bones. This article presents an example where this is performed by using the segmentation of the exterior of the dental and facial bone structures. However, the inner structures of the mouth and sinuses are probably of equal clinical importance. Although not presented in this article, the algorithm was also tried on the interior surfaces of the mouth. These tests suggest that by readjusting the parameters, it is possible to segment the interior with similar accuracy as we now report for the exterior. Ideally, the pipeline should be configured to segment both the inner and outer surfaces simultaneously, exploiting the obvious similarities in shape, orientation and location of

Table 4 Averaged segmentation result images in Study 1 (S1), Study 2 (S2) and both (all) combined

Set	Mean (mm)	Std (mm)	RMS (mm)	Coverage (%)
S1	0.52	1.14	1.26	89
S2	0.49	0.93	1.05	94
all	0.50	1.01	1.13	92

RMS, root mean squares; Std, standard deviation.

these structures. This is an obvious topic for further research. Segmenting the maxillary sinuses is a rather different kind of challenge for which methods exist.^{19,20} Although not tested, we believe that our algorithm has potential for segmenting the sinuses also.

A key feature of the algorithm is the use of parameterization to define the topology to define some properties of the surface to represent the target. The choice of using a two-dimensional, third-degree polynomial fitted to the facial skeleton was made for its resemblance to the setting of traditional panoramic imaging. This also led to the very intuitive relation between the parameterized and real-world volume spaces and a very simple mathematic

description. In other applications (unknown pose, for example), this would be a limitation since it results in a poor ability to capture edges parallel to axial planes. In our case, moving to full three-dimensional surface parameterization would have resulted in significant and complicating changes with only potentially marginal improvements to the segmentation accuracy. We have tested the algorithm successfully in segmenting another target scanned with another modality by using genus 0-type closed surface for parametrization.²

A short execution time is a key property of any algorithm intended for clinical use. We believe that an algorithm with a runtime of 1 min, and less when

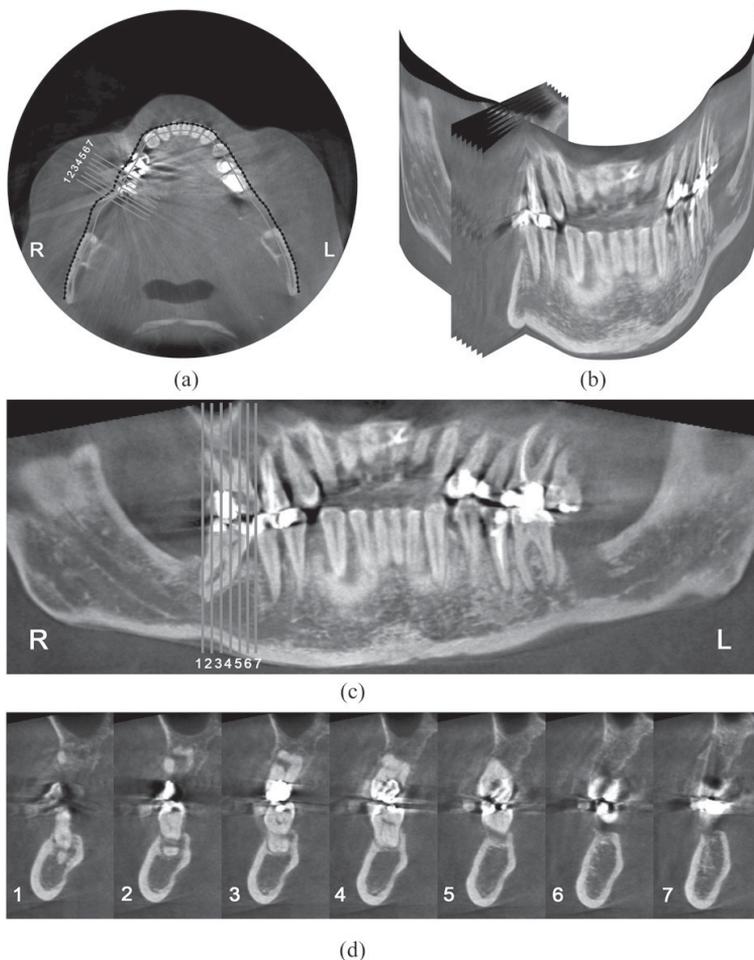


Figure 11 An example of using the segmentation result to reconstruct panoramic and cross-sectional slices: a line contour (grey line) to mark the centre of a layer used in panoramic reconstruction is drawn on a slice of interest (a). The contour is computed by moving the segmentation contour (black dotted line) inwards to a fixed distance, here 4 mm (a). Seven cross-sections (grey lines) are placed on the right molar (a). A scene including the reconstruction of the panoramic and cross-sectional slices in (a) is rendered in three dimensions showing the curvature of the panoramic layer (b). The panoramic reconstruction laid on a plane shows the molar roots and the mandibular canal (c). The cross-sectional slices marked on (a) show the extent of the roots relative to the mandibular canal (d). This example was created using Image 26 in our data set. L, left; R, right.

implemented efficiently on modern hardware, would provide a dentist or oral surgeon enough time to ensure a smooth workflow, for example, for inspecting the processed images during a single patient visit. Although only those algorithms that can be implemented efficiently were chosen for our pipeline, the presented implementation was intended to demonstrate only the feasibility of our approach and was not optimized for performance. The current main bottlenecks of the pipeline, the thin plate spline computation and Gaussian + Sobel filtering, could be significantly improved with parallelization. The nature of those tasks readily permits this. The same applies to the computation of the voxel energies. We believe that exploiting the power of modern graphical processing unit computing would bring the execution time to a fraction of the current version.

Conclusions

In conclusion, we developed an algorithm for segmenting the exterior of the facial skeleton from CBCT

References

- De Vos W, Casselman J, Swennen GR. Cone-beam computerized tomography (CBCT) imaging of the oral and maxillofacial region: a systematic review of the literature. *Int J Oral Maxillofac Surg* 2009; **38**: 609–25. doi: <https://doi.org/10.1016/j.ijom.2009.02.028>
- Antila K, Lijja M, Kalke M, Lotjonen J. Automatic extraction of mandibular bone geometry for anatomy-based sythenization of radiographs. *Conf Proc IEEE Eng Med Biol Soc* 2008; **2008**: 490–3. doi: <https://doi.org/10.1109/IEMBS.2008.4649197>
- Dreiseidler T, Alarabi N, Ritter L, Rothamel D, Scheer M, Zöller JE, et al. A comparison of multislice computerized tomography, cone-beam computerized tomography, and single photon emission computerized tomography for the assessment of bone invasion by oral malignancies. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2011; **112**: 367–74. doi: <https://doi.org/10.1016/j.tripleo.2011.04.001>
- Monteiro BM, Nobrega Filho DS, Lopes Pde M, de Sales MA. Impact of image filters and observations parameters in CBCT for identification of mandibular osteolytic lesions. *Int J Dent* 2012; **2012**: 239306. doi: <https://doi.org/10.1155/2012/239306>
- Closmann JJ, Schmidt BL. The use of cone beam computed tomography as an aid in evaluating and treatment planning for mandibular cancer. *J Oral Maxillofacial Surg* 2007; **65**: 766–71. doi: <https://doi.org/10.1016/j.joms.2005.12.053>
- Schulze R, Heil U, Gross D, Bruellmann DD, Dranischnikow E, Schwanecke U, et al. Artefacts in CBCT: a review. *Dentomaxillofac Radiol* 2011; **40**: 265–73. doi: <https://doi.org/10.1259/dmfr/30642039>
- Lamecker H, Zachow S, Wittmers A, Weber B, Hege H, Elsholtz B, et al. Automatic segmentation of mandibles in low-dose CT-data. *Int J Comput Assist Radiol Surg* 2006; **1**: 393–5.
- Cootes TF, Taylor CJ, Cooper DH, Graham J. Active shape models—their training and application. *Comput Vis Image Und* 1995; **61**: 38–59. doi: <https://doi.org/10.1006/cviu.1995.1004>
- Rueda S, Gil JA, Pichery R, Alcañiz M. Automatic segmentation of jaw tissues in CT using active appearance models and semi-automatic landmarking. *Med Image Comput Comput Assist Interv* 2006; **9**(Pt 1): 167–74.
- Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. *IEEE Trans Pattern Anal Mach Intell* 2001; **23**: 681–5. doi: <https://doi.org/10.1109/34.927467>
- Kainmueller D, Lamecker H, Seim H, Zinser M, Zachow S. Automatic extraction of mandibular nerve and bone from cone-beam CT data. *Med Image Comput Comput Assist Interv* 2009; **12** (Pt 2): 76–83.
- Wang L, Chen KC, Gao Y, Shi F, Liao S, Li G, et al. Automated bone segmentation from dental CBCT images using patch-based sparse representation and convex optimization. *Med Phys* 2014; **41**: 043503. doi: <https://doi.org/10.1118/1.4868455>
- Collins DL, Neelin P, Peters TM, Evans AC. Automatic 3D inter-subject registration of MR volumetric data in standardized Talairach space. *J Comput Assist Tomogr* 1994; **18**: 192–205. doi: <https://doi.org/10.1097/00004728-199403000-00005>
- Lötjönen JM, Wolz R, Koikkalainen JR, Thurfjell L, Waldemar G, Soiminen H, et al; Alzheimer's Disease Neuroimaging Initiative. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage* 2010; **49**: 2352–65. doi: <https://doi.org/10.1016/j.neuroimage.2009.10.026>
- Mazziotta JC, Toga AW, Evans A, Fox P, Lancaster J. A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM). *Neuroimage* 1995; **2**: 89–101. doi: <https://doi.org/10.1006/nimg.1995.1012>
- Pianykh OS. Analytically tractable case of fuzzy c-means clustering. *Pattern Recogn* 2006; **39**: 35–46. doi: <https://doi.org/10.1016/j.patcog.2005.06.005>
- Brechbühler C, Gerig G, Kübler O. Parametrization of closed surfaces for 3-D shape description. *Comput Vis Image Und* 1995; **61**: 154–70. doi: <https://doi.org/10.1006/cviu.1995.1013>
- Bookstein FL. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans Pattern Anal Mach Intell* 1989; **11**: 567–85. doi: <https://doi.org/10.1109/34.24792>
- Descoteaux M, Audette M, Chinzei K, Siddiqi K. Bone enhancement filtering: application to sinus bone segmentation and simulation of pituitary surgery. *Med Image Comput Comput Assist Interv* 2005; **8**(Pt 1): 9–16.
- Last C, Winkelbach S, Wahl FM, Eichhorn KWG, Bootz F. A model-based approach to the segmentation of nasal cavity and paranasal sinus boundaries. *Pattern Recognition* 2010; vol. 6376 of Lecture Notes in Computer Science. pp. 333–42.

PUBLICATION
V

**Automatic segmentation for detecting uterine fibroid regions treated with
MR-guided high intensity focused ultrasound (MR-HIFU)**

Kari Antila, Heikki J. Nieminen, Roberto Blanco Sequeiros and Gösta Ehnholm

Medical Physics, 41(7), p. 073502
<https://doi.org/10.1118/1.4881319>

Publication reprinted with the permission of the copyright holders.

Automatic segmentation for detecting uterine fibroid regions treated with MR-guided high intensity focused ultrasound (MR-HIFU)

Kari Antila, Heikki J. Nieminen, Roberto Blanco Sequeiros, and Gösta Ehnholm

Citation: Medical Physics **41**, 073502 (2014); doi: 10.1118/1.4881319

View online: <http://dx.doi.org/10.1118/1.4881319>

View Table of Contents: <http://scitation.aip.org/content/aapm/journal/medphys/41/7?ver=pdfcov>

Published by the American Association of Physicists in Medicine

Articles you may be interested in

Dynamic T2-mapping during magnetic resonance guided high intensity focused ultrasound ablation of bone marrow

AIP Conf. Proc. **1503**, 222 (2012); 10.1063/1.4769948

Robotic active positioning for magnetic resonance-guided high-intensity focused ultrasound

AIP Conf. Proc. **1503**, 217 (2012); 10.1063/1.4769947

Dual-modality image guided high intensity focused ultrasound device design for prostate cancer: A numerical study

AIP Conf. Proc. **1481**, 131 (2012); 10.1063/1.4757323

Comparison between diffusion-weighted imaging, T₂-weighted, and postcontrast T₁-weighted imaging after MR-guided, high intensity, focused ultrasound treatment of uterine leiomyomata: Preliminary results

Med. Phys. **37**, 4768 (2010); 10.1118/1.3475940

A method for MRI guidance of intercostal high intensity focused ultrasound ablation in the liver

Med. Phys. **37**, 2533 (2010); 10.1118/1.3413996

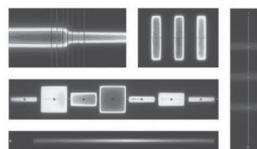
View this article online for free at scitation.aip.org



RITG148[†]

Automated QA for helical tomotherapy

- Y-jaw divergence/beam centering
- Y-jaw/gantry rotation plane alignment
- Treatment field centering
- Image quality tests (Cheese Phantom)
- Built in trending and reporting with RITtrend
- MLC alignment test
- Couch translation/gantry rotation
- Laser localization



Visit us at AAPM
Booth #401

Automatic segmentation for detecting uterine fibroid regions treated with MR-guided high intensity focused ultrasound (MR-HIFU)

Kari Antila^{a)}

VTT Technical Research Centre of Finland, Tampere, FI-33200 Tampere, Finland

Heikki J. Nieminen

MR-Therapy, Philips Healthcare, FI-01511 Vantaa, Finland and Department of Physics, University of Helsinki, FI-00014, Helsinki, Finland

Roberto Blanco Sequeiros

South Western Finland Imaging Centre, Turku University Hospital and Turku University, FI-20521, Turku, Finland

Gösta Ehnholm

MR-Therapy, Philips Healthcare, FI-01511 Vantaa, Finland

(Received 29 January 2014; revised 18 April 2014; accepted for publication 15 May 2014; published 12 June 2014)

Purpose: Up to 25% of women suffer from uterine fibroids (UF) that cause infertility, pain, and discomfort. MR-guided high intensity focused ultrasound (MR-HIFU) is an emerging technique for noninvasive, computer-guided thermal ablation of UFs. The volume of induced necrosis is a predictor of the success of the treatment. However, accurate volume assessment by hand can be time consuming, and quick tools produce biased results. Therefore, fast and reliable tools are required in order to estimate the technical treatment outcome during the therapy event so as to predict symptom relief.

Methods: A novel technique has been developed for the segmentation and volume assessment of the treated region. Conventional algorithms typically require user interaction or *a priori* knowledge of the target. The developed algorithm exploits the treatment plan, the coordinates of the intended ablation, for fully automatic segmentation with no user input.

Results: A good similarity to an expert-segmented manual reference was achieved (Dice similarity coefficient = 0.880 ± 0.074). The average automatic segmentation time was 1.6 ± 0.7 min per patient against an order of tens of minutes when done manually.

Conclusions: The results suggest that the segmentation algorithm developed, requiring no user input, provides a feasible and practical approach for the automatic evaluation of the boundary and volume of the HIFU-treated region. © 2014 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4881319>]

Key words: uterine fibroid, high-intensity focused ultrasound, image segmentation, treatment volume

1. INTRODUCTION

Up to 25% of women suffer from uterine fibroids (UF), which are benign muscle tumors.¹ Magnetic resonance-guided high-intensity focused ultrasound (MR-HIFU) has proved to be a promising method for treating these tumors noninvasively.² The ultimate benefit is that it requires no surgical incisions, resulting in short recovery times and a patient-friendly safety profile. The treatment typically leads to symptom improvement, and is able reduce the size of the fibroid.

The volume of thermally (HIFU) ablated fibroid tissue predicts the success of alleviating the tumor-related symptoms.³ Determining the treated fibroid region with conventional measurement tools, such as approximation with a manually drawn ellipsoid, may be straightforward; however, the ablated tissue as seen in the postoperative images may be very irregular or complex in shape.⁴ Therefore, characterization of the treated region, especially volume measurement, may be strongly limited and biased by conventional, oversimplifying approaches. Precise manual delineation of the treated volume is subjective and too laborious and time-consuming to be performed

in a clinical setting. Therefore, segmentation methods, potentially automated, would contribute towards a more precise, more rapid, and less burdensome assessment of the treatment outcome.

A wealth of segmentation methods has been developed for detecting various targets in MR images; semiautomatic techniques rely on user input,⁵ while others utilize statistical information such as intensity, shape, size, and location of known data.^{6,7} Available processing time, required precision, and the amount of pre-existing data affect what methods can be considered. Given that the treated region may have a complicated and unpredictable shape, the classic statistical or model-based segmentation methods utilizing prior data may not be feasible. A semiautomatic method of region growing has previously been applied to problem.⁸ In region growing, the operator identifies the target by placing a number of seed points that are grown to larger regions by iteratively adding neighboring voxels with similar properties to the region. Although region growing-based techniques are able to label consistent areas with similar properties, the results are generally highly dependent on the seed selection and similarity parameter. There are

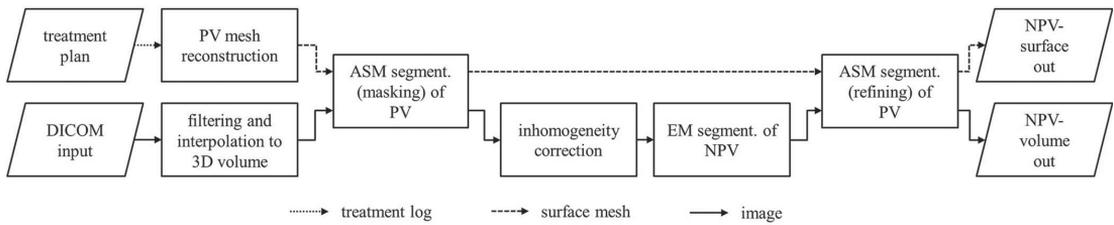


FIG. 1. A schematic presentation of the automatic segmentation pipeline. The pipeline took postoperative, contrast-enhanced MR slices of the fibroid in DICOM and the treatment log containing sizes and orientations of the sonication cells (focal points) of HIFU thermal ablation as inputs. Images were preprocessed, stacked and interpolated to 3D volumes, and a surface model for the fibroid was created based on the sonication cell data. The surface model was deformed around the PV with an ASM (Active Shape Model)-type segmentation algorithm. The voxels inside the deformed surface were contained in a mask, inside which the voxels in the PV (the untreated tissue) and the NPV (the treated tissue) were classified by an EM (Expectation Maximization) segmentation algorithm. The surface model was refitted to the EM-segmented volume, and the NPV regions output in binary volume for quantification and as a surface mesh for visual evaluation.

currently no techniques that would automatically, precisely, and expediently segment the treated region from contrast-enhanced MR images during or following a HIFU treatment.

The aim of this study was to develop and validate an algorithm for automatic segmentation of treated regions in an MR image stack taken immediately after treatment. Since MR-HIFU is a computer-controlled technique in which the HIFU-beam is focused on predefined locations called treatment cells, the recorded coordinates of the cells could be exploited in the segmentation in order to ease or remove the requirement for user input.⁹ We, therefore, utilized the size, shape, location, and orientation of each treatment cell (i.e., the volume intended to be thermally ablated by HIFU) as *a priori* information in the development of the algorithm.

2. MATERIALS AND METHODS

2.A. Subjects and MR imaging

Thirty-one subjects were scanned at four different clinics (sites 1–4, $n = 16, 2, 3, 10$) after receiving HIFU treatment (MR-HIFU for uterine fibroid therapy, Philips MR-therapy, Vantaa, Finland) in a previous clinical trial.¹⁰ The trial was authorized by local boards of ethics. T1-weighted images (THRIVE, TFE3D with SPAIR and SENSE, TR/TE 6.6 ms/3.2 ms, FA 10°, IR delay 90 ms, SENSE 1.5, Scan time 4:40 min, pixel size 0.48×0.48 mm, 3 mm slice thickness) were acquired after administration of 0.1 mmol per body kg of Gadopentetate Dimeglumine. The contrast agent visualized the nonperfused volume (NPV), i.e., fibroid tissue considered to have been ablated by HIFU in contrast to the perfused volume (PV), the fibroid tissue with remaining blood circulation.

Treatment logs including the sizes, locations, orientations, and completion of the HIFU sonication of each treatment cell were obtained. Data from 2 out of 31 subjects were omitted because of a failure to induce NPV due to technical or other reasons. Preoperative MR-images were available from only one clinic (site 1, $n = 16$) and were used to confirm the lack of pre-existing necrosis. The lack of pre-existing necrosis was an exclusion criterion for the trial, so for the rest of the subjects it was assumed.

2.B. Segmentation

The segmentation algorithm was implemented as a pipeline of several consecutive steps (Fig. 1). The pipeline began by reading postoperative, contrast-enhanced MRI image slices, and the treatment log file. The image slices were filtered with a Gaussian kernel ($\sigma = 0.5$ mm) for suppressing noise, and linearly interpolated to a three-dimensional volume of 1 mm^3 isotropic voxels. The sizes, locations, and orientations of the sonication cells in the image coordinates were read from the treatment log; however, the cells whose sonication was reported as incomplete by the HIFU system were discarded. The surfaces of the cluster of cells were meshed with triangles [Fig. 2(a)], and a convex hull was formed on the mesh nodes of the cluster by three-dimensional Delaunay

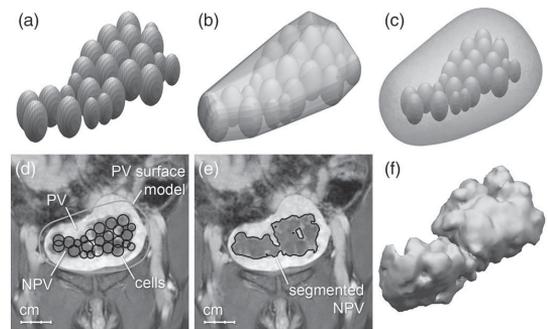


FIG. 2. The ASM segmentation method required an initial surface model having roughly the right size, shape and location for the target. The model was formed by first meshing a number of ellipsoids representing the sonication cells in the treatment log (a). Next, a convex hull was formed on the nodes of the meshed cell surfaces (b). The hull mesh was enlarged towards the presumed outer border of the PV so as to cover the approximate safety margins of the sonication plan, resampled for even triangulation and smoothed for a more rounded appearance (c). PV is the round object at the center of the image containing the more complex-shaped and darker object, NPV (d). When overlaid on a contrast-enhanced MRI image, this initial PV surface model (gray contour) and cell surfaces (black, round objects) should have approximately covered the PV and NPV, respectively. NPV was segmented with the EM algorithm (e). The segmented NPV regions were also output as a surface mesh (f).

triangulation [Fig. 2(b)]. The outer surface of the hull was remeshed for equally spaced triangles and enlarged by moving every mesh node to the direction of the out-pointing surface normal by 10 mm [Fig. 2(c)]. The mesh was enlarged since the treatment plan has a safety margin in order to prevent damage to the surrounding tissues. Thus the treatment cell cluster is always smaller than the fibroid. The enlarged mesh was subsequently smoothed for a more rounded appearance by moving the mesh vertices iteratively towards the average location of the neighboring vertices in ten fixed iterations.

The purpose of the mesh was approximately to contain the fibroid tissue and the treated tissue. These are essentially the PV and the NPV. These were segmented from the image with an active-shape-model (ASM) type algorithm.¹¹ The algorithm worked in its classic form by moving the mesh nodes along their normal vectors towards the maximum of image edge gradients, and the goodness-of-fit to the surface was determined as the sum gradients of voxels in current mesh node locations according to the energy function

$$E_t = \sum_i \bar{g}_x * \bar{n}_i^T, \quad (1)$$

where t is iteration number at iteration, \bar{g}_x is the gradient vector at voxel coordinate x , and \bar{n}_i is the normal vector for mesh node i . The strength of the gradients was normalized to a maximum of 1. As a slight modification to the classic ASM, the gradients were multiplied with the corresponding mesh node normal vectors \bar{n}_i [Eq. (1)]. This gives the strongest response to edges having the same local orientation as the surface being deformed. The energy was computed on voxels x along the normal vector within ± 25 mm from the node i and moved towards the maximum. No internal energy term regularizing the shape of the PV surface was used. The fitting process was run iteratively until 95% of the nodes had converged or a fixed limit of 15 iterations was reached. Once converged, the surface should have contained all of the NPV and the surrounding areas of the PV. The final result was not affected if a part of the PV was left out as long as all of the NPV was included.

The ASM-segmented surface was used to form a mask that separates the NPV and parts or all of the PV from the rest of the image. The inside of the surface-masked region was corrected for inhomogeneity, which were often present in scans with a large field-of-view (FOV), using the N4ITK algorithm.¹² The voxels inside the mask were segmented to two tissue classes representing PV and NPV with an expectation-maximization (EM) algorithm.^{13,14} The EM algorithm fitted a mixture of a predefined number of Gaussian distributions on the intensity histogram of the voxels inside the mask. The version of the EM algorithm used also utilized local information in the form of Markov Random Fields (MRF).^{14,15} MRF added a probability component that depends on the classifications of neighboring voxels. This gave a smoother and less noisy result when compared to solely intensity-based EM-segmentation. The MRF model used in this paper¹⁴ did not require any user-defined parameters to be set. All voxels were labeled as PV or NPV; class PV was considered to be more probable in the brighter and NPV in the darker side of the spectrum.

The EM-segmentation typically results in a label map on which the PV-labeled region contains all NPV-labeled voxels. However, it is not uncommon for NVP-labeled regions to penetrate the PV shell. This may happen, where the HIFU beam was targeted close to border of the fibroid, where dark local image artifacts were present, or where large blood vessels penetrated the PV. This may result in leaks—darker-than-PV areas of other tissues surrounding the PV but still contained by the ASM-segmented surface—being misclassified as NPV. To cap these leaks, the PV surface was refitted to the PV/NPV class label image, as given by the EM-algorithm, with the same ASM segmentation method as used in deforming the original surface. This time, the surface was refitted specifically to the binary object of voxels labeled as PV. As a result, the surface was tightly deformed around the PV-labeled object still containing the NPV.

The 6-connected regions of NPV-labeled voxels inside the refitted mask were analyzed, and those isolated regions with volumes smaller than the volume of the smallest sonication cell in the treatment log, or those that have no overlap with any of the sonication cells in the log, were discarded. The label image of NPV was retained for quantitative volumetric analysis and the NPV-objects were meshed with the Delaunay triangulation algorithm for visual inspection.

The majority of the algorithm was implemented with MathWorks MATLAB 2012b (Natick, MA). Some of performance critical parts of the PV segmentation, and all of the EM segmentations were implemented in C/C++. N4ITK was used as part of the Insight Registration Toolkit (ITK), version 4.2.¹⁶

2.C. Validation

The proposed segmentation results were compared to a reference segmentation drawn manually by an expert (H.J.N.) and verified by an experienced radiologist (R.B.S.). NPVs were drawn on the coronal MR slices of the contrast-enhanced images. NPVs resulting from the treatment were verified by comparing images taken pre- and post-treatment during the manual segmentation if available. The reference segmentations were drawn using Osirix software (version 3.7.1 32-bit) with the help of a region growing tool, followed by manual slice-by-slice inspection and manual correction of the NPV boundary.

The automatic and reference segmentations were compared with the Dice similarity coefficient (DSC).¹⁷ In addition, the volumes of the regions where the compared regions disagreed were computed (Fig. 3). False negative rate (FNR) is the number of voxels included in the manual reference but not in the automatically segmented region, and false positive rate (FPR) is the opposite. Both are relative to the total number of voxels in the manual reference. In addition, the manual reference regions were meshed with the Delaunay triangulation to three-dimensional surfaces, and average point-to-surface distances (in mm) between the automatic and reference both ways were computed.

The data set was randomly split into two groups; the first half was used in teaching the algorithm ($T, n = 14$), while the

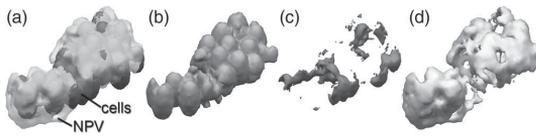


FIG. 3. Visual comparison of the sonication cell cluster and the NPV surfaces. A cell cluster surface is shown with the resulting NPV surface (a). Both are meshed from binary image form, and in this case the surfaces match and align well. The intersection of voxels covered in both objects is called the true positives (b). It is not uncommon for HIFU therapy to leave some small areas covered by the cells to end up untreated (c). When the NPV and sonication cell regions are compared, the relative number of untreated voxels to all voxels inside the cells is called the false negative rate (FNR). In many cases, and especially when larger sonication cell sizes are used, the NPV induced by the thermal ablation is larger than the cell cluster (d). The number of these voxels relative to all voxels inside the cells is the false positive rate (FPR). In this figure, the NPV is compared to the sonication cells when in Sec. 3 the automatically segmented NPV is compared to the manual reference.

other half was used for validation ($V, n = 15$). It was further ensured that at least one sample from each clinic was included in the teaching and validation group. This split was carried out prior to the development of the algorithm, and the division was kept throughout the study. The algorithm was taught by testing the parameter combinations, namely the weight of the MRF term in EM-segmentation, and several pipeline configurations by trial and error for the best average DSC for the teaching set. The validation set was analyzed only after the parameter and pipeline configurations for the algorithm were determined.

The proposed algorithm was benchmarked against three well known segmentation methods: Otsu's thresholding,¹⁸ region growing,⁵ and Fuzzy C-means clustering.¹⁹ While none of these fully automatic methods could replace the entire proposed pipeline, they are suitable for comparisons of performance in the NPV segmentation with the EM-algorithm (Fig. 1). All the methods chosen are capable of binary segmentation, have no or few parameters to be set and are widely used in image processing. The distance-to-mean parameter for region growing was estimated by exhaustive trial and error segmentations of the T group and set to 8,6% from the maximum intensity.

The entire method was validated using a regular Intel i7 (4 core, 2.6 GHz), Windows 7 laptop workstation.

3. RESULTS

Visual examples of the worst, average, and best segmentation results in DSC are shown in Fig. 4. A Bland-Altman plot of the difference in volumes measured from the automatic and reference segmentations [Fig. 5(a)] or the samples by sample results (Table I) revealed no major outliers with only one sample outside of the mean $\pm 1.96 \times$ S.D. boundary [Fig. 5(a)]. The difference between the NPV volumes measured from the two segmentations was 0.17 ± 6.77 ml (mean \pm S.D.) for all samples, 0.50 ± 8.26 ml for the T and -0.16 ± 5.16 ml V group. An average DSC of 0.880 ± 0.074 for all samples between the compared NPVs was achieved (5B, Table II). Our proposed method concluded in better Dice sim-

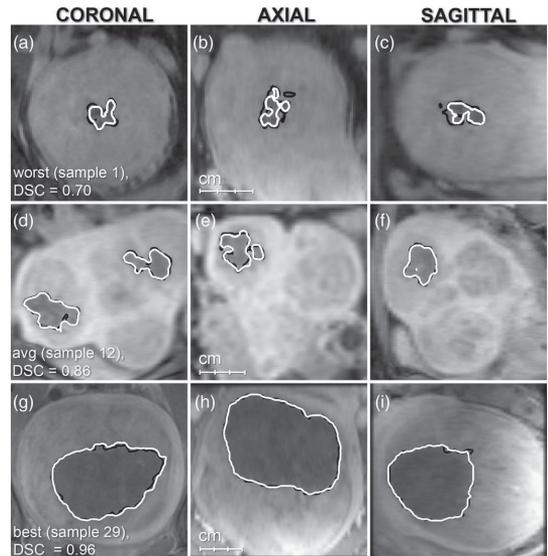


FIG. 4. Visual examples of the worst (a)–(c), average (d) and (e), and best (g)–(i) segmentation results as measured with the Dice similarity coefficient (DSC). The hand-segmented NPV contour is presented by black, and the automatically segmented NPV contour by the white line. The segmentation accuracy (in terms of DSC) was typically improved as the treatment volume got larger. When small volumes were treated, the deposited thermal energy was subjected to thermal diffusion and perfusion, resulting in imperfect necrosis (a)–(c). Also, as the DSC measure is relative to the sizes of the objects compared, even small absolute deviations in shape give poor values in small objects.

ilarity when compared to the three reference methods: Otsu 0.788 ± 0.177 , region growing 0.755 ± 0.238 and Fuzzy C-means 0.772 ± 0.193 (Table II). The first seven samples arranged in ascending (worst to best) order according to DSC also showed the highest FNR [Fig. 5(b)]; i.e., with small NPVs the automatic algorithm underestimated the size. On the other hand, the samples with low DSC did not appear to have a severe shape deviation. This can be seen from the average distances from the nodes of the automatically segmented surface to the reference surface [auto. to ref., Fig. 5(c)]. The plot shows that the average auto. to ref. distances of samples 1–7 did not differ from the rest of the set. The 2+ mm spikes in Fig. 5(c) (sample 2, ref. to auto., samples 8 and 20 auto to ref.) are explained by protuberances or small isolated surfaces not present in the surface to be compared. Figure 5(d) shows that the lowest DSCs are measured from the smallest segmented volumes. The average DSC of NPVs for all samples above 30 ml was 0.921 ± 0.031 (Otsu 0.876 ± 0.069 , region growing 0.869 ± 0.074 , Fuzzy C-means 0.871 ± 0.070). Figure 5(e) shows that volume of the cell cluster in the treatment log was somewhat correlated with the volume of the NPVs in smaller regions, but not on the larger ones.

The average execution time on the Intel i7 (Windows 7 workstation) was 1.6 ± 0.7 min per subject. The time depends heavily on the size of the 3D volume and the surface area of the objects to be segmented. Approximately 85% of the total

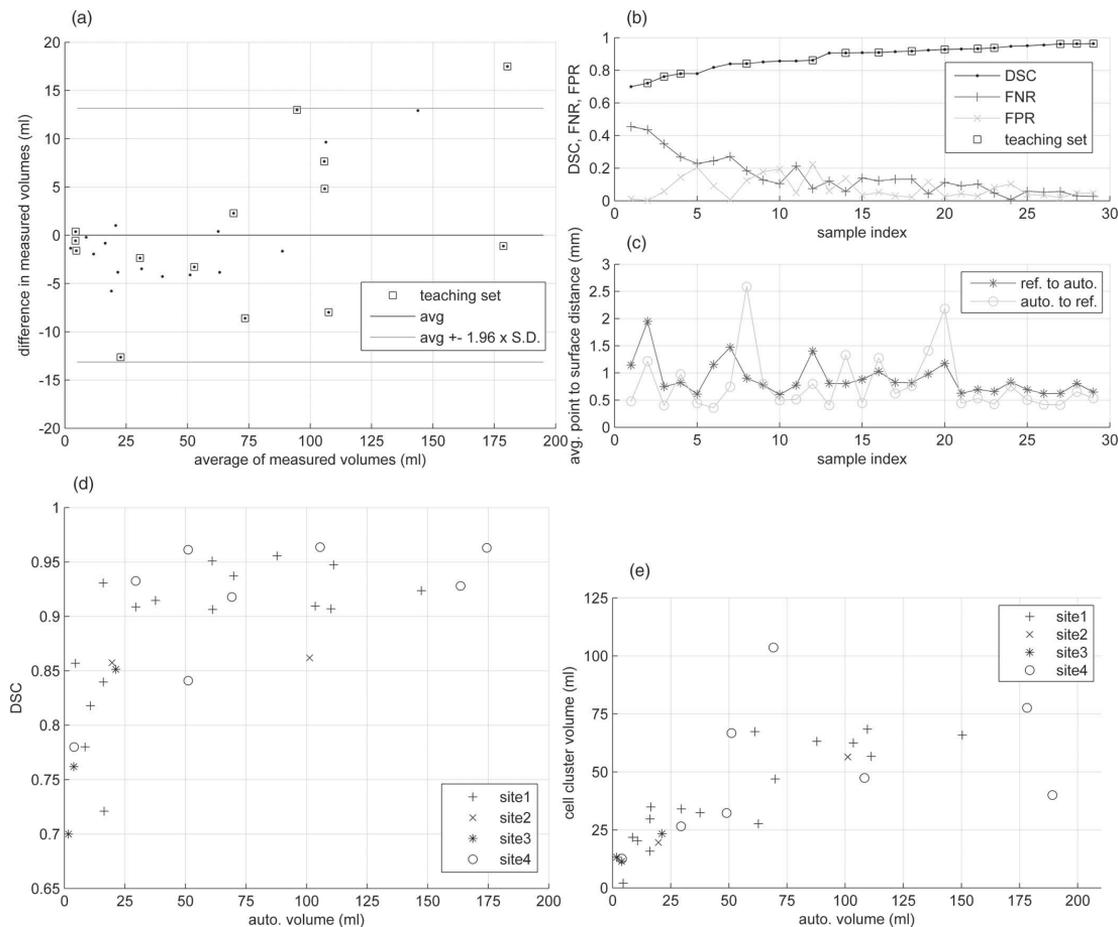


FIG. 5. An excellent agreement between the volumes of the automatically segmented NPV and the manual reference was reached as shown by Bland-Altman plot (a). The plot shows a slight tendency of the automatic algorithm to underestimate the volume of the smaller NPV regions. For all samples, the average difference between the measured volumes is very small, 0.17 ml. Samples belonging to the teaching set are indicated on the DSC plot with a square. Figure 5(b) plots DSC, FNR, and FPR values of data sets in ascending order according to the DSC. The worst samples (1–7) according to the DSC also show the highest FNR; thus, the automatic method underestimated the size of NPV in these samples. Point to surface distances from mesh nodes to the surface measured both ways show roughly an average accuracy of 1 mm with some peaks (c). The peaks indicate protuberances or small isolated objects not found in the surface to be compared to. Figure 5(d) shows further evidence of poorer DSC performance in small treatment volumes, especially site 3 ($n = 3$). NPV volumes of 30 ml and above gave an average accuracy of DSC = 0.921. Figure 5(e) shows the tendency of the automatic algorithm in producing smaller NPVs than the treatment cell volume below 30 ml. From 30 to 80 ml, there may be a vague correlation between the volumes. NPVs sized over 80 ml did not seem to correlate with cell cluster size. This is probably due to the secondary effects of the thermal ablation present with higher doses. In Fig. 5(e), the treatment cell volumes are plotted against the automatically segmented NPV volumes, but the results were similar with the manual reference.

pipeline execution time was spent on the two ASM segmentation steps (Fig. 1). The tested implementation was configured for verifying the feasibility of the method, and thus was not optimized for speed. For example, the segmentation pipeline was implemented single-threaded and included a lot of writing and reading of the intermediate results on the hard drive. The time measured does not include the preprocessing or interpolation of the slice stack to 3D volume that typically takes less than ten seconds, most of which is disk I/O. The reference segmentations by hand took in the order of tens of minutes per patient.

4. DISCUSSION

Based on the results, the automatic segmentation algorithm developed was able to give consistent results with a very good match to the reference boundary drawn manually by an expert (Figs. 4 and 5). The likeness of the automatic and hand-segmented NPV regions as measured with the Dice similarity coefficient was found to be high (0.880 ± 0.074 for all samples, 0.885 ± 0.0780 for the T and 0.876 ± 0.072 for the V groups) [Fig. 5(b), Table II]. Since the volume of the necrosed tissue induced by the thermal ablation is the main indicator of

TABLE I. Sample-specific segmentation data with samples ordered according to DSC, as also plotted in Fig. 5(b).

Sample index	Site	Set	DSC	Auto. volume (ml)	Ref. volume (ml)	FNR	FPR
1	3	V	0.70	3	2	0.45	0.01
2	1	T	0.72	29	16	0.44	0.00
3	3	T	0.76	5	4	0.35	0.06
4	4	T	0.78	5	4	0.27	0.14
5	1	V	0.78	9	9	0.23	0.21
6	1	V	0.82	13	11	0.24	0.09
7	1	V	0.84	22	16	0.27	0.01
8	4	T	0.84	54	51	0.18	0.12
9	3	V	0.85	20	21	0.13	0.18
10	1	T	0.86	4	5	0.10	0.19
11	2	V	0.86	24	20	0.21	0.05
12	2	T	0.86	88	101	0.01	0.22
13	1	V	0.90	65	61	0.12	0.06
14	1	T	0.90	102	110	0.06	0.14
15	1	V	0.91	33	30	0.14	0.03
16	1	T	0.91	112	104	0.12	0.05
17	1	V	0.91	42	38	0.13	0.03
18	4	T	0.92	78	69	0.13	0.02
19	1	V	0.92	137	147	0.04	0.12
20	4	T	0.93	179	164	0.11	0.03
21	1	V	0.93	17	16	0.09	0.04
22	4	T	0.93	32	29	0.10	0.03
23	1	T	0.94	68	70	0.05	0.08
24	1	V	0.95	102	111	0.01	0.10
25	1	V	0.95	62	61	0.06	0.04
26	1	V	0.96	89	88	0.05	0.04
27	4	V	0.96	53	51	0.06	0.02
28	4	T	0.96	172	174	0.03	0.05
29	4	T	0.96	103	106	0.03	0.05

the success of the treatment, the result is excellent considering the very complex shapes of the NPV. The computation time even in its current nonoptimized form was reasonable with a conventional laptop computer and thus feasible to be implemented for a clinical environment.

In all of the samples with $DSC < 0.8$, the automatic segmentation method underestimated the volume of NPV [Fig. 5(b), Table I]. This is the result of a small intended treatment volume; when only a small portion of the fibroid is treated, the deposition of thermal energy is diffused and perfused away, resulting in nonperfect necrosis. The use of small sonication cells not only increases the risk of heat diffusion, but also results in NPV with blurry borders in contrast-enhanced MR images taken immediately after the treatment. This makes the classification of voxels on the PV/NPV border difficult, also for a human observer.

The success of the automatic segmentation method relies on two steps: ASM segmentation of the outer border of the PV in order to find the mask that contains all NPV and EM-segmentation to detect the voxels with NPV inside the mask. ASM segmentation relies partially on the treatment log to become properly initialized. Also, the postprocessing steps of the segmentation of NVP rely somewhat on the treatment log. Based on the available data, the treatment log seems to match reasonably well the NPV as seen in the postoperative images. There is still a risk of patient or intestine movement prior or during the postoperative scan, which could misalign the fibroid from the location of the sonication cells fixed to the MR coordinate system. The ASM-segmentation method used was able to tolerate the slight movement when necessary. However, large subject or fibroid movement should be compensated by other means. One practical way to solve this would be to use conventional tools for registering the pre- and

TABLE II. Averaged results of the proposed and the three reference methods for teaching (T) and validation (V) sample groups and all samples (all) pooled together.

Set	Measurement	Proposed EM-based method		Otsu		Region growing		Fuzzy C-means	
		Avg	Std	Avg	Std	Avg	Std	Avg	Std
T	DSC	0.885	0.080	0.801	0.185	0.808	0.226	0.794	0.189
V	DSC	0.876	0.072	0.776	0.174	0.705	0.246	0.751	0.202
All	DSC	0.880	0.074	0.788	0.177	0.755	0.238	0.772	0.193
T	FNR	0.14	0.13	0.04	0.06	0.19	0.12	0.04	0.06
V	FNR	0.15	0.11	0.03	0.03	0.26	0.21	0.02	0.03
All	FNR	0.15	0.12	0.04	0.04	0.22	0.11	0.03	0.04
T	FPR	0.072	0.063	0.64	1.08	1.93	7.04	0.69	1.18
V	FPR	0.080	0.066	0.71	0.87	3.18	11.2	0.91	1.18
All	FPR	0.076	0.064	0.68	0.96	2.57	9.27	0.81	1.17
T	Ref. to auto. dist.	0.93 (mm)	0.36 (mm)	0.99 (mm)	0.51 (mm)	1.38 (mm)	0.87 (mm)	1.01 (mm)	0.51 (mm)
V	Ref. to auto. dist.	0.85 (mm)	0.25 (mm)	1.01 (mm)	0.48 (mm)	1.86 (mm)	1.41 (mm)	1.08 (mm)	0.57 (mm)
All	Ref. to auto. dist.	0.89 (mm)	0.31 (mm)	1.00 (mm)	0.48 (mm)	1.63 (mm)	1.18 (mm)	1.05 (mm)	0.53 (mm)
T	Auto. to ref. dist.	1.01 (mm)	0.67 (mm)	3.59 (mm)	3.94 (mm)	2.82 (mm)	5.23 (mm)	4.00 (mm)	3.94 (mm)
V	Auto. to ref. dist.	0.59 (mm)	0.27 (mm)	3.22 (mm)	4.70 (mm)	3.39 (mm)	6.42 (mm)	3.92 (mm)	5.12 (mm)
All	Auto. to ref. dist.	0.79 (mm)	0.54 (mm)	3.40 (mm)	4.28 (mm)	3.11 (mm)	5.78 (mm)	3.96 (mm)	4.51 (mm)

postoperative image stacks. The ASM-segmentation method does not require precise registration and, thus, even quick manual approximation such as dragging the surface model to a new location on the treatment console would probably work.

The EM-segmentation is critical in identifying the voxels belonging to NPV. The decision whether a voxel in the mask encapsulated by the ASM segmentation is classified as PV or NPV relies mostly on probabilities derived from the intensity distribution. An MRF probability component was used in order to achieve a smoother and less noisy result. The weight of the MRF component in the probability estimation was kept rather low, since this appeared to give better results for the smaller NPVs in the teaching set. For the larger NPVs, the weight of MRF did not make much difference. While MRF is useful in bringing local information to the classification, it probably cannot compensate cases where the intensity distributions are severely skewed or overlap extensively. This may happen for example if the inhomogeneity artifacts are not compensated successfully earlier in the pipeline. We also tested the idea of making an *a priori* probability map based on the voxels inside the cells of the treatment log for the EM algorithm. This works well with larger sonication cell clusters, but may give false estimates with smaller ones. Using such a probability map did not improve overall results, so it was eventually left out.

The proposed EM-segmentation method for NPV was tested against three widely used segmentation methods: Otsu's thresholding region growing and Fuzzy C-means clustering. Our method was found to give more accurate results through the NPV size range, especially on the smaller NPVs. The difference in similarity in larger, 30+ ml NPVs was smaller but still clear. An interesting finding is that the EM method slightly underestimates the NPV (larger FNR than FPR in Table II), all reference methods overestimate the NPV (smaller FNR than FPR in Table II). This is probably because of the MRF component used with the EM method.

The region growing method was found very sensitive to the selection of the distance parameter. The difference in results of the T (DSC = 0.808) and V (DSC = 0.705) sample groups for region growing was the largest for all tested methods. The proposed method performs almost equally on T (DSC = 0.885) and V (DSC = 0.876) groups. This suggests that the success of the method is not dependent on the selection of the single parameter, the weight of the MRF component. We believe that the method would have performed similarly with a smaller training group. Importantly the smaller and more blurry NPVs were represented in the training group; therefore, methods capable of segmenting also these challenging samples were selected for the pipeline.

While the technique is intended for segmenting nonperfused volumes in uterine fibroids imaged using MR, we foresee that this approach may be usable for verifying treatment outcome in other clinical indications; i.e., segmentation of treatment volume based on label or label-free contrast could be used in order to verify outcome of MR-HIFU therapy in

treatment of pathologies in organs such as the liver, prostate or breast.

5. CONCLUSIONS

In conclusion, we have developed a fully automatic segmentation algorithm for predicting the treatment outcome during MR-HIFU therapy of uterine fibroids. We have found the method fast and, precise and thus feasible to be utilized in a clinical context.

ACKNOWLEDGMENTS

The study was supported by the SalWe Research Program for IMO (Tekes—the Finnish Funding Agency for Technology and Innovation grant 648/10). The salary of Heikki J. Nieminen was partially covered by the Academy of Finland (253579).

^{a)} Author to whom correspondence should be addressed. Electronic mail: kari.antila@vtt.fi

¹E. A. Stewart, "Uterine fibroids," *Lancet* **357**, 293–298 (2001).

²H. Ruhnke, T. Eckey, M. K. Bohlmann, M. P. Beldoch, A. Neumann, A. Agic, J. Hagele, K. Diedrich, J. Barkhausen, and P. Hunold, "MR-guided HIFU treatment of symptomatic uterine fibroids using novel feedback-regulated volumetric ablation: Effectiveness and clinical practice," *Rofo* **184**, 983–991 (2013).

³Z. M. Lénárd, N. J. McDannold, F. M. Fennessy, E. A. Stewart, F. A. Jolesz, K. Hynynen, and C. M. Tempny, "Uterine leiomyomas: MR imaging-guided focused ultrasound surgery—Imaging predictors of success1," *Radiology* **249**, 187–194 (2008).

⁴K. R. Gorny, D. A. Woodrum, D. L. Brown, T. L. Henrichsen, A. L. Weaver, K. K. Amrami, N. J. Hangiandreou, H. A. Edmonson, E. V. Bouwsma, E. A. Stewart, B. S. Gostout, D. A. Ehman, and G. K. Hesley, "Magnetic resonance-guided focused ultrasound of uterine leiomyomas: Review of a 12-month outcome of 130 clinical patients," *J. Vasc. Interv. Radiol.* **22**, 857–864 (2011).

⁵W. K. Pratt, "Image segmentation," in *Digital Image Processing: PIKS Scientific Inside*, 4th ed. (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2006).

⁶T. Heimann and H.-P. Meinzer, "Statistical shape models for 3D medical image segmentation: A review," *Med. Image Anal.* **13**, 543–563 (2009).

⁷J. M. Lötjönen, R. Wolz, J. R. Koikkalainen, L. Thurfjell, G. Waldemar, H. Soininen, and D. Rueckert, "Fast and robust multi-atlas segmentation of brain magnetic resonance images," *NeuroImage* **49**, 2352–2365 (2010).

⁸C. Militello, S. Vitabile, G. Russo, G. Candiano, C. Gagliardo, M. Midiri, and M. C. Gilardi, "A semi-automatic multi-seed region-growing approach for uterine fibroids segmentation in MRgFUS treatment," in *2013 Seventh International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS)*, Taichung (IEEE Conference publications, Taichung, Taiwan, 2013), pp. 176–182.

⁹H. J. Nieminen, M. O. K. Köhler, and M. Häkkinen, "Medical imaging system, computer-implemented method, and computer program product for identifying a treated region in a medical image," U.S. patent application 20130208961 A1 (Filing date Oct 13, 2011).

¹⁰M. Voogt, H. Trillaud, Y. Kim, W. T. M. Mali, J. Barkhausen, L. Bartels, R. Deckers, N. Frulio, H. Rhim, and H. Lim, "Volumetric feedback ablation of uterine fibroids using magnetic resonance-guided high intensity focused ultrasound therapy," *Eur. Radiol.* **22**, 411–417 (2012).

¹¹T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—their training and application," *Comput. Vis. Image Understanding* **61**, 38–59 (1995).

¹²N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4ITK: Improved N3 bias correction," *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).

¹³T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.* **13**, 47–60 (1996).

- ¹⁴K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain," *IEEE Trans. Med. Imaging* **18**, 897–908 (1999).
- ¹⁵E. Salli, H. J. Aronen, S. Savolainen, A. Korvenoja, and A. Visa, "Contextual clustering for analysis of functional MRI data," *IEEE Trans. Med. Imaging* **20**, 403–414 (2001).
- ¹⁶National Library of Medicine, Insight Segmentation and Registration Toolkit (ITK) (14 November 2013) (available URL: <http://www.itk.org>).
- ¹⁷K. H. Zou, S. K. Warfield, A. Bharatha, C. Tempny, M. R. Kaus, S. J. Haker, W. M. Wells III, F. A. Jolesz, and R. Kikinis, "Statistical validation of image segmentation quality based on a spatial overlap index: Scientific reports," *Acad. Radiol.* **11**, 178–189 (2004).
- ¹⁸N. Otsu, "A threshold selection method from gray-level histograms," *Systems, Man and Cybernetics, IEEE Transactions* **9**(1), 62–66 (1979).
- ¹⁹D. Dunn and W. E. Higgins, "Optimal Gabor filters for texture segmentation," *IEEE Trans. Image Processing* **4**, 947–964 (1995).

