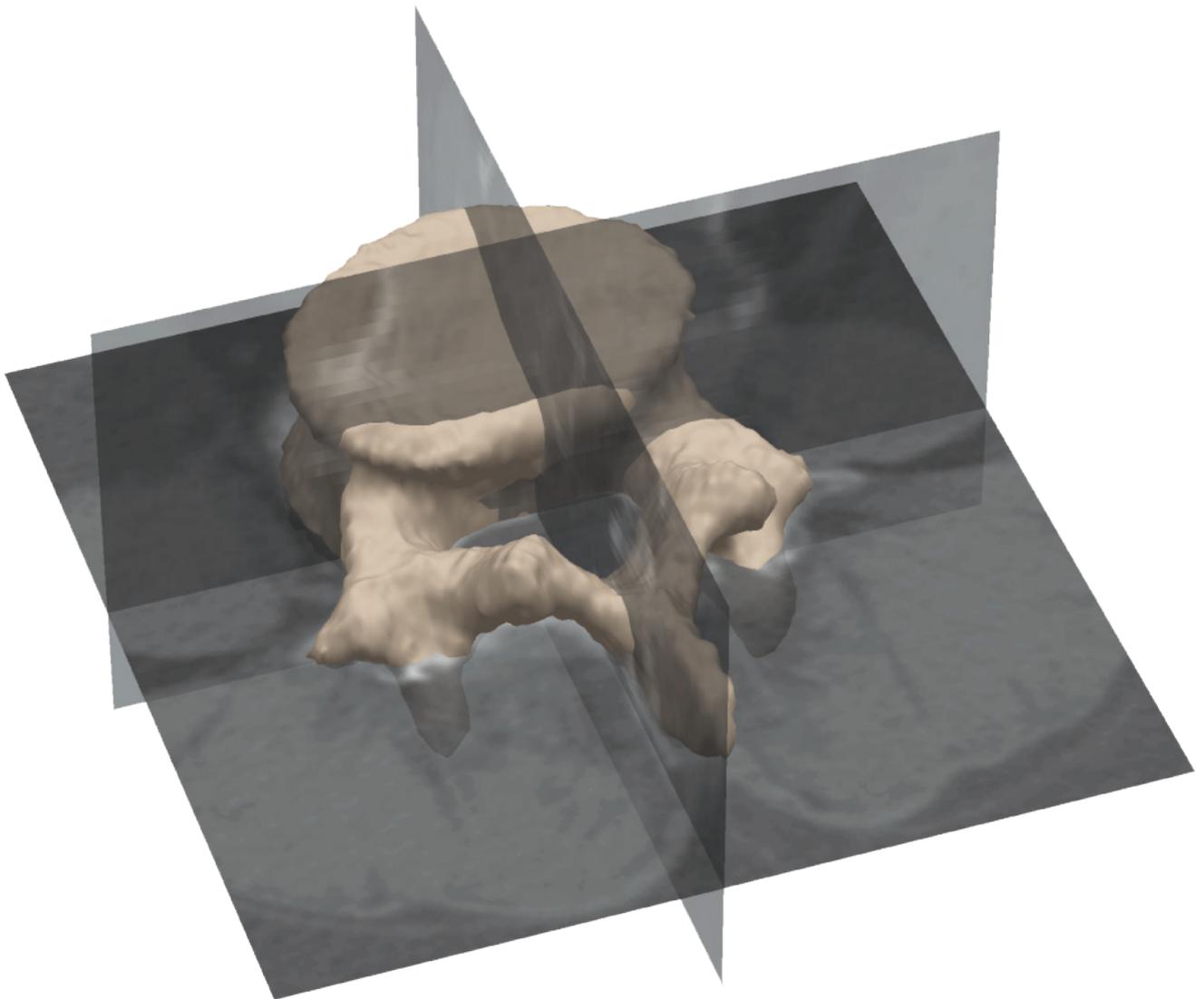


LEVEL-SET APPEARANCE MODELING

FOR SEGMENTATION OF ANATOMICAL STRUCTURES
IN 3-D MEDICAL IMAGES



ULRIK LANDBERG STEPHANSEN
FOURTH SEMESTER, MASTER – 2012
BIOMEDICAL ENGINEERING AND INFORMATICS



Title:

Level-set appearance modeling for segmentation of anatomical structures in 3-D medical images

Subject:

Applied biomedical engineering and informatics

Project period:

4th semester, master, BMEI
February 1st–June 1st 2012

Project group:

12gr1076

Author:

Ulrik Landberg Stephansen

Supervisors:

Lasse Riis Østergaard
Anne Sofie Korsager

Print run: 4

Report page count: 83

Appendices page count: 8

ABSTRACT:

Medical imaging is being increasingly used, and the demand for automatic segmentation of structures of interest grow.

In this work a model-based segmentation method is implemented: An active appearance model based on principal component analysis with a level-set representation of shape is utilized in an iterative algorithm for segmentation of 3-D images. The automatic segmentation algorithm incorporates prior knowledge to predict how to correct model and pose parameters in order to achieve a better fit of the model to the target image.

The segmentation method is tested on 42 prostate MR images and 27 CT images of the L4 vertebra in a leave-one-out cross-validation framework. The automatic segmentations are compared to manual reference segmentations. A median Dice kappa of 0.81 is achieved for both structures.

The algorithm performs similar to previously described methods, but in some cases it fails to determine the correct size of the prostate. Also the appearance model is not large enough to fully segment the vertebral processes. The algorithm is sensitive to the initial location of the average model in the target image. The active appearance model presented can be applied on any imaging modality and any structure of interest if the shape of the structure is not too variable.

Preface

This report is my masters thesis concluding my education as Master of Science in Biomedical Engineering and Informatics at Aalborg University. The project has been done in collaboration with McConnell Brain Imaging Center at Montreal Neurological Institute, McGill University, Canada and the Department of Medical Physics at Aalborg Hospital, Denmark.

I would like to thank D. Louis Collins from McConnell Brain Imaging Center for providing manually segmented vertebra images, as well as Jesper Carl from the Department of Medical Physics at Aalborg Hospital for exporting manually annotated prostate images from the hospital's RIS/PACS.

I would also like to thank Vladimir Fonov from McConnell Brain Imaging Center for helping out with building the MINC tool kit and related tools.

Aalborg University, June 1st 2012

Ulrik Landberg Stephansen

Contents

Preface	i
1 Introduction	1
I Background	
2 The prostate	5
2.1 Prostate cancer	6
2.2 Magnetic resonance imaging of the prostate	8
2.3 Prior work in prostate segmentation	10
3 The vertebra	13
3.1 Lumbar spondylodosis	14
3.2 Computed tomography of the vertebrae	15
3.3 Prior work in vertebra segmentation	15
II Methods	
4 Data and tools	19
4.1 Programming framework	20
5 Appearance modeling	21
5.1 Shape representation	21
5.2 Texture	23
5.3 PCA-based models	24
5.4 Level set appearance model	24
6 Segmentation	27
6.1 Transformation between reference frames	27
6.2 Learning parameter adjustments	28
6.3 Iterative segmentation algorithm	30
7 Image pre-processing	33
7.1 Common reference	33
7.2 Registration	33
7.3 Image volumes	36
7.4 Manual segmentations	37
7.5 Preprocessing pipeline implementation	38
8 Performance evaluation	41

III Results

9	Prostate segmentation	47
9.1	Appearance model	47
9.2	Parameter adjustment predictions	50
9.3	Segmentation performance	51
10	Vertebra segmentation	57
10.1	Appearance model	57
10.2	Parameter adjustment predictions	60
10.3	Segmentation performance	60

IV Discussion

11	Discussion	69
11.1	Conclusion and directions for future work	74

	Bibliography	75
--	---------------------	-----------

Appendices

A	Representing 3-D pose	I
B	Pipeline implementation	III

Introduction

A major challenge in the field of medical image analysis is to solve the need for automatic segmentation of anatomical structures in 3-D images. Segmentations are used both for diagnosis and treatment. In diagnostics the segmentation of organs and other internal structures can be used to identify e.g. abnormal changes in size and shape. An example is the segmentation of the cerebral cortex to diagnose neurodegenerative diseases [Eskildsen & Østergaard, 2006].

Segmentation is also used for radiotherapy planning, in which the target volume and organs at risk are delineated to deliver a high dose to the cancerous tissue while minimizing the exposure of sensitive normal tissue. Another example is the emerging field of image-guided surgery, where pre- and/or per-operative images are used for guiding the surgeon during surgery, thus improving surgery outcome and reducing the risk of side effects.

The limited resolution and quality of medical images pose a challenge to segmentation. Typical clinical 3-D images from X-ray computed tomography (CT) or magnetic resonance imaging (MRI) have voxel sizes in the order of 1–3 mm. This results in a blurring of the image due to the partial volume effect, a result of each voxel representing an average of the tissue within the voxel. The image quality can also be reduced by noise, patient movement, or artifacts introduced as a result of the imaging modality. In CT imaging high density structures, e.g. implanted metal objects, can introduce streak artifacts seen as radial beams emerging from the high-density objects. In MRI intensity non-uniformities can arise as a result of radiofrequency field inhomogeneities and patient anatomy, resulting in otherwise homogeneous tissue having non-uniform intensity values [Erasmus et al., 2004]. Similarly magnetic field inhomogeneities result in image distortion.

These artifacts pose a challenge when developing automatic segmentation techniques. As medical imaging is being increasingly used, the demand for automatic techniques to reduce the manual labor associated with segmentation grows.

Current segmentation methods span from the simpler region growing methods [Gonzalez & Woods, 2008] to model-based techniques, e.g. active contours [Kass et al., 1988] and active shape/appearance models [Cootes & Taylor, 2004]. In region growing techniques the segmentation expands from a seed point into nearby voxels of sim-

ilar properties, but these techniques often fail when facing low contrast or some of the artifacts mentioned above, especially partial volume effects leading to leakage, or beam hardening causing the propagation to stop prematurely. In a more sophisticated technique the image is matched to a set of atlases, or manually labeled images, to produce the segmentation. However, some structures in the human body, e.g. the sulci of the brain, are so complex in their variability between humans that they can not be sufficiently represented by a set of atlases [Caunce & Taylor, 1999].

Model-based techniques aim at describing this variability statistically to generate new shapes from a description of the mean shape and the shape variability. However, the training data for model-based methods are typically provided as manually identified anatomical landmarks combined in a point distribution model to represent the shape of the object of interest. A major problem with landmark-based approaches is that landmarks from different images in the training set have to correspond with each other in an anatomically meaningful way for the statistical point distribution model to make sense.

Another approach to shape representation is to use a signed distance map [Tsai et al., 2003]. In this method the zero level-set of the distance map represents the surface of the object of interest. The level-set shape representation has been combined with several model-based segmentation methods, more recently the Active Appearance Modeling (AAM) method [Hu & Collins, 2007]. This level-set AAM method was applied to segment the ventricles of the brain from MRI, but can be readily applied to other anatomical structures and imaging modalities with the constraint that the shape must not vary too much.

The aim of this project is to implement and evaluate the level-set AAM segmentation method on other anatomical structures and imaging modalities. Training data is available for the prostate (MRI) and the L4 vertebra (CT), and these will be the subject of the evaluation. The goal is to build a statistical appearance model of the anatomical structures, not including possible pathologies (tumors, etc.). Pathologies cause a larger variation in shape and appearance, requiring a much larger training set to build a statistical model, if it is at all possible. Vertebral segmentation can be applied for image-guided surgery in spondylodesis, and prostate segmentation can be applied for radiotherapy planning in prostate cancer treatment.

PART



Background

The prostate

The prostate is a small round muscular gland encompassing the urethra inferior to the bladder (Fig. 2.1). The organ has a diameter of roughly 4 cm. It secretes prostatic fluid, which comprise 20–30% of semen and has an antibiotic effect among others, which protects the urethra from infections. The prostatic fluid is ejected into the urethra by peristaltic contractions of the smooth muscle in the prostate. [Martini, 2006]

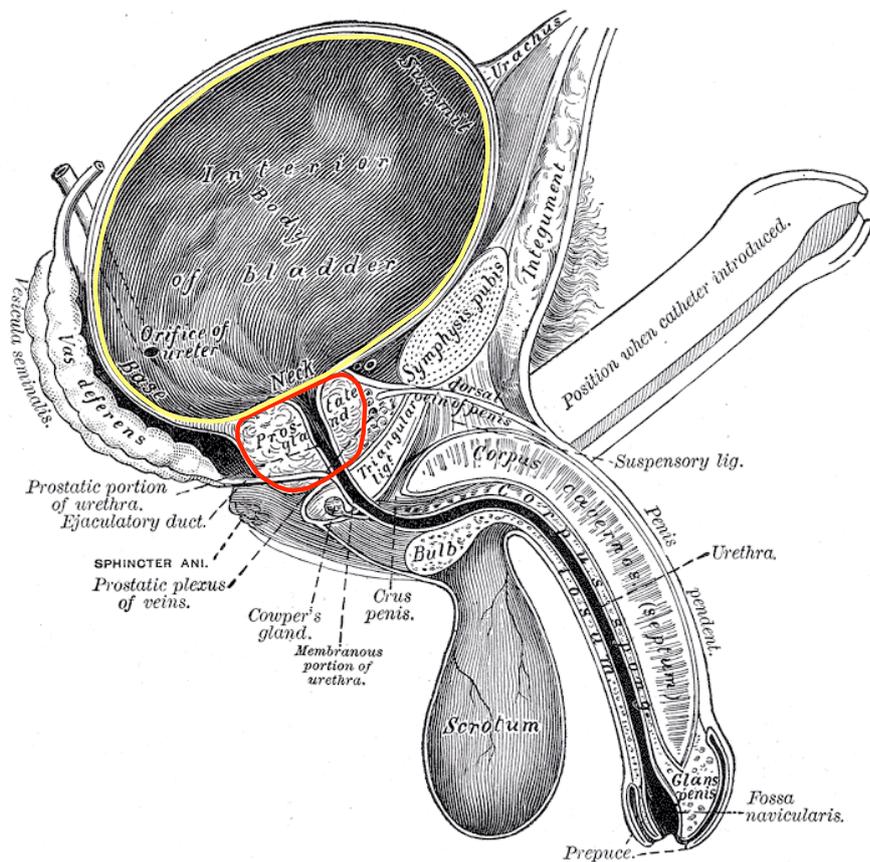


Figure 2.1: Male anatomy, sagittal view. The prostate (outlined in red) is located inferior to the bladder (outlined in yellow) and envelops the urethra. The seminal vesicles are located posterior to the prostate and bladder. [Gray, 1918].

2.1 Prostate cancer

Prostate cancer originates in the secretory glands of the prostate [Martini, 2006], and is the most frequent type of cancer among Danish males [Danish Health and Medicines Authority, 2009]. 14% of cancer-related deaths among males are caused by prostate cancer, or 4% of all deaths among males. This makes it the second most frequent cause of cancer related deaths among males, with lung cancer being the most frequent. [Danish Health and Medicines Authority, 2009]

In Denmark the incidence has increased from 91.1 per 100,000 males in 2000 to 136.9 per 100,000 in 2007, which is a 50% increase. Part of the rise in incidence can be explained by better diagnostics and more focus on prostate cancer. In the same time-frame the prevalence increased from 336.3 per 100,000 males to 616.0 per 100,000, an 83% increase. However, the mortality rate is almost unchanged with 53.2 per 100,000 males in 2000 decreasing to 49.3 per 100,000 in 2007. [Danish Health and Medicines Authority, 2009]

The risk factors for prostate cancer are not very well understood and subject to research, but age and family history plays an important role. It is estimated that 5–10% of prostate cancer cases are hereditary [Brasso, 2007]. Most new cases are seen among 65–69 year-olds, and prostate cancer is rare before the age of 55 [Danish Health and Medicines Authority, 2009].

2.1.1 Diagnosis

The main procedures for diagnosing prostate cancer are palpation, blood test for increased prostate-specific antigen (PSA) level, and ultrasound [Martini, 2006].

By doing palpation through the rectal wall a cancerous prostate appears as a hard, irregular nodule. This procedure is known as digital rectal exam (DRE), a relatively easy procedure which is performed routinely on patients with possible prostate cancer. The positive predictive value is, however, rather low, but combining positive findings during palpation with increased PSA levels the precision of the test is improved. Prostate-specific antigen is secreted by the prostate, and prostate cancer often causes increased levels of PSA in the bloodstream. A test for increased PSA level is, however, not a good indicator for prostate cancer on its own as there is a number of benign causes of increased PSA level. PSA is not recommended for screening purposes in Denmark. [Jønler & Pedersen, 2007]

Transrectal prostatic ultrasound (TRUS) is a more precise diagnostic modality during which biopsies of the tumor and prostate can be taken. Using TRUS more detailed information about tumor location and size can be obtained. Typically around ten biopsies are taken for subsequent histological studies. [Jønler & Pedersen, 2007]

2.1.2 Treatment

The choice of treatment modality depends on the prognosis and tumor state, but to some extent also on the patient's preference. The prognosis depends on PSA level, Gleason score (a tumor grading system based on histological appearance), comorbidity, and the patient's age [Jakobsen et al., 2007].

For patients with localized cancer, no metastases, a good prognosis, and no other health problems present, the typical treatment choice is active monitoring, in which the development of the tumor is followed closely, but no actual treatment is initiated [Jakobsen et al., 2007; Martini, 2006]. Patients with an expected lifespan of more than 10–15 years and no metastases can be offered surgical removal of the prostate (prostatectomy) or radiotherapy, whereas patients with a shorter expected lifespan are offered medical endocrine therapy [Jakobsen et al., 2007; Jønler & Pedersen, 2007].

In Denmark the number of males undergoing medical therapy has increased from 74.3 per 100,000 males in 2003 to 136.4 per 100,000 in 2007. During the same time-frame the number of males undergoing radiotherapy has increased from 32.9 per 100,000 to 57.4 per 100,000 males. The number of males undergoing prostatectomy has increased from 4.1 per 100,000 males in 2000 to 22.8 per 100,000 in 2007. [Danish Health and Medicines Authority, 2009]

2.1.2.1 Prostatectomy

Prostatectomy is surgical removal of the prostate and possibly the seminal vesicles and some surrounding tissue depending on tumor size and location. The treatment is offered to patients with an expected lifespan of more than 10 years. Prostatectomy can be done using open surgical techniques or laparoscopic techniques, possibly robot assisted. The laparoscopic technique shortens admission time, but does not otherwise reduce side effects or improve mortality rate. [Jakobsen et al., 2007]

The most frequent side effect is loss of erectile function, as the two pudendal nerves innervating the penis runs around the inferior part of the prostate. Depending on the extent of the tumor one or both of these nerves can be preserved during surgery. The nerve-preserving surgical technique, as well as post-surgery medical prophylaxis, have been found to improve recovery of erectile function. Whether to use the nerve-preserving technique or not depends on the erectile function before surgery, because the erectile function deteriorates with age. [Borre et al., 2008]

Urinary incontinence is another frequent side effect, but most patients recover within 3–6 months. Pre-surgical training of the pelvic diaphragm is found to improve recovery time. Ideally less than 5% should experience urinary incontinence after 12 months. [Jakobsen et al., 2007]

2.1.2.2 Radiotherapy

Radiotherapy can be given in the form of external beam radiotherapy (EBRT) or brachytherapy (BT), in which radioactive beads are implanted in the prostate. Radiotherapy is offered to patients under 75 years of age having localized cancer [Jakobsen et al., 2007]. The treatment has been reported to be as effective as prostatectomy [Fokdal & Høyer, 2005].

Radiotherapy treatment of prostate cancer is the third most frequent type of radiotherapy in Denmark, with 23,273 radiation fractions given and 1411 patients being treated with prostate cancer using radiotherapy in 2007. The number of patients with prostate cancer being treated using radiotherapy increased 20% from 2003 to 2007, and it is expected that half of newly diagnosed patients will be treated using radiotherapy in the future. [Jensen et al., 2009]

The prostate can move up to 10 mm in each direction depending on bladder filling and other factors [Jakobsen et al., 2007]. This and other issues make it necessary to add a margin to the radiation target volume, causing surrounding normal tissue to receive a relatively high dose of radiation. The side effects from radiotherapy are caused by this exposure of normal tissue to radiation, and include anorectal, bladder, and sexual dysfunctions [Fokdal & Høyer, 2005]. Late side effects occur months to years after the treatment, and the damage to normal tissue is irreversible. The severity of side effects depends on radiation dose and volume of normal tissue included in the target volume. [Fokdal & Høyer, 2005]

In conformal radiotherapy the target volume is delineated in 3-D images, thus determining the size, shape and location of the prostate. This geometrical information is used for planning the shape, direction and intensity of the radiation beams to reduce the target volume and exposure of normal tissue [Fokdal & Høyer, 2005; Jakobsen et al., 2007] leading to fewer and less severe side effects [Dearnaley et al., 1999; Vordermark et al., 2003]. The prostate size and shape are currently determined by manually outlining the prostate in each image slice, a tedious and error-prone task.

2.2 Magnetic resonance imaging of the prostate

3-D images are used to plan the conformal radiotherapy in order to deliver a high dose to the tumor and a low dose to the surrounding tissue. Traditionally CT is used, because radiotherapists need the electron density values calculated from the CT image for dose planning. However, MRI offer better soft tissue contrast using T2-weighted recording, thus the anatomy of the prostate and tumor margins can be better visualized. The surrounding soft tissue has the same attenuation coefficient as the prostate in CT images, and combined with the partial volume effect this makes it virtually impossible

to determine which slice represents the lower apex of the prostate in CT images. These factors make it easier to delineate the prostate from MR images, resulting in a smaller target volume and decreasing the inter-observer variance [Villeirs & Meerleer, 2007].

The delineation of the prostate from MRI can be transferred to CT images for use in radiotherapy planning. Current methods for MRI to CT registration span from manually identifying anatomical landmarks in each image volume and calculating the transformation between each landmark set to fully automatic registration techniques. In a method combining the two approaches a stent was placed in the prostatic urethra and used as fiducial marker for landmark-based registration [Østergaard et al., 2010]. The registration was then refined using an automatic method based on mutual information [Collins et al., 1994].

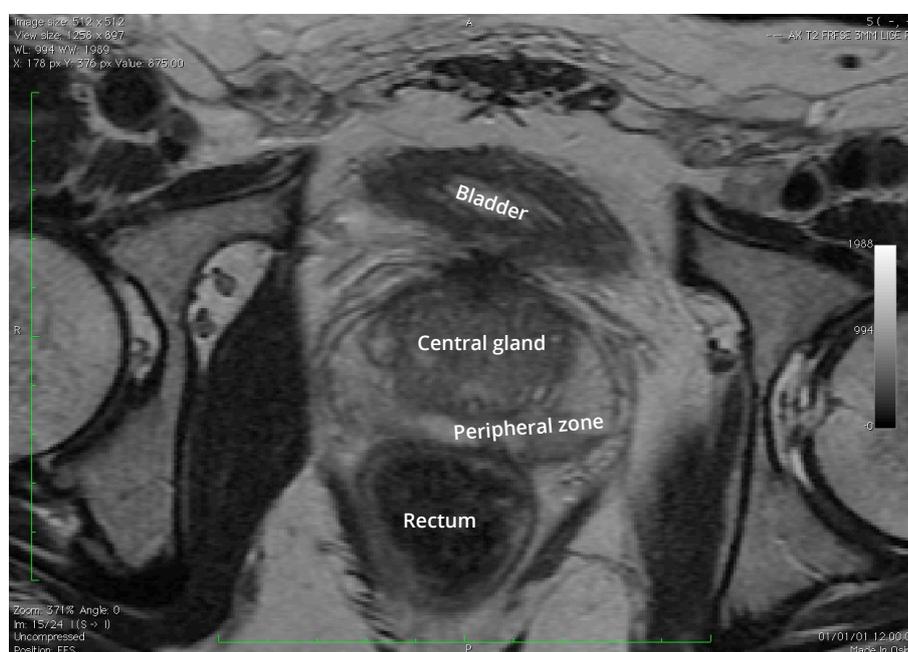


Figure 2.2: *The prostate and surrounding structures as appearing on MRI. The prostate consists of the nodular central gland and the higher signal intensity peripheral zone. Image is prostate cancer patient 5 scan 1 from the data set.*

In MR images the prostate appears as two regions; the central gland and the peripheral zone (Fig. 2.2). This is especially the case in elderly men, as the peripheral zone expands with age. The central gland appears as a nodular region with high signal intensity nodules representing glandular secretions and low signal intensity regions representing muscular and fibrous tissue. The peripheral zone lies posterior to the central zone and has high signal intensity representing the glandular tissue. The prostate is encapsulated by the prostatic capsule, which appears as a dark rim around the prostate. [Villeirs & Meerleer, 2007]

A set of manually labeled atlas images was registered to an unknown image using non-rigid registration, and the labels were merged using different weighing or selection schemes. This method was able to automatically segment the prostate with a Dice kappa of 0.85 which was close to the inter observer variability of 0.87 [Klein et al., 2008].

Other studies have investigated the use of model-based segmentation techniques for fully automatic prostate segmentation. One study implementing a landmark-based shape model using active contours for deforming the model was able to segment the prostate on MRI with a mean Jaccard similarity of 0.78 ± 0.05 [Pasquier et al., 2007].

A landmark-based appearance model using Haar wavelets for extracting texture features was able to segment the prostate on MRI with a mean Dice kappa of 0.88 ± 0.06 and a mean absolute distance of 3.97 ± 2.74 mm [Ghose et al., 2010].

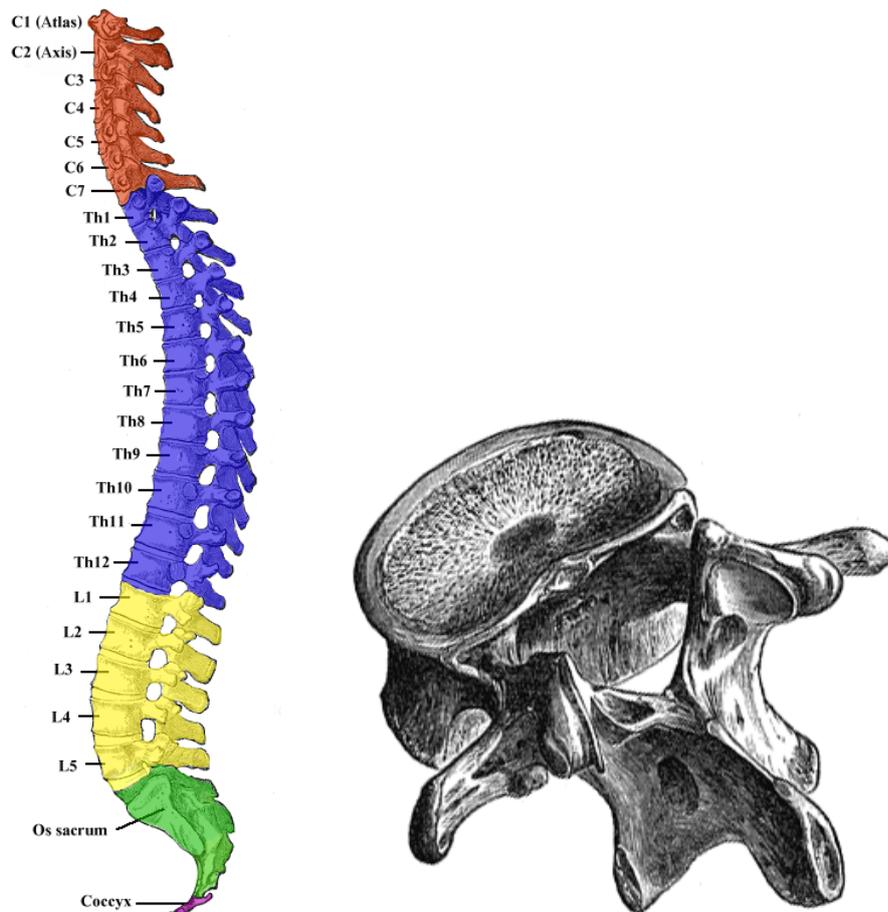
The main problem with landmark-based shape and appearance models is that landmark points in different training volumes have to anatomically correspond to each other in some meaningful way. The landmark representation of shape also puts some constraints on the type and variability of the structures represented by the model [Cootes & Taylor, 2004].

A different modeling approach used signed distance maps for representing the shape in a level-set framework, and used Chan-Vese active contours for curve evolution [Tsai et al., 2001]. The technique was also demonstrated on the problem of prostate segmentation [Tsai et al., 2004]. A recent study applied the level-set shape representation in a shape model to achieve simultaneous multi-modal segmentation of the prostate in MRI and CT images [Chowdhury et al., 2012]. Segmentation was done by fitting the shape model to binary images extracted from MRI and CT using probabilistic atlases and a random forest classifier based on Haar wavelet and gradient features. A mean Dice kappa of 0.83 ± 0.056 and a mean absolute distance of 2.6 mm were achieved for segmentation of the prostate from MRI [Chowdhury et al., 2012].

The level-set shape model has also been incorporated with the active appearance model to yield an appearance model which does not depend on landmarks [Hu & Collins, 2007], and that method is applied in this project.

The vertebra

The main functions of the vertebral column are to support the upper body and to protect the spinal cord. The vertebral column consists of seven cervical vertebrae supporting the head, twelve thoracic vertebrae supporting the ribs, five lumbar vertebrae supporting the upper body, five fused sacral vertebral segments between the hips, and three to five coccygeal segments constituting the coccyx (Fig. 3.1(a)). The vertebrae are named sequentially according to segment and position starting from the head.



(a) The five segments of the vertebral column. (b) A lumbar vertebra. Anterior side is in the top left corner, posterior side in the bottom right.

Figure 3.1: Anatomy of the human vertebral column and lumbar vertebrae [Reworked from Gray, 1918].

Except for C1 and C2 the vertebrae consists of a vertebral body from which the pedicles protrude on the posterior side (Fig. 3.1(b)). The bodies of adjacent vertebrae are separated by an intervertebral disk. The two pedicles merge forming the laminae and the vertebral arch.

The posterior surface of the vertebral body and the anterior surface of the laminae form the spinal canal holding the spinal cord. The spinous process arise from the posterior side of the vertebral arch, and the two transverse processes arise from the sides. In the thoracic vertebrae the transverse processes articulate with the ribs.

The superior and inferior articulate processes arise from the laminae, and form facet joints with the adjacent vertebrae. The orientation of the facet joint determines the range of movements permitted. The cervical facet joints have an almost horizontal orientation favoring rotation, while the lumbar facet joints have a more sagittal orientation favoring flexion and extension. [Hosten & Liebig, 2002]

3.1 Lumbar spondylodesis

In lumbar spondylodesis (spinal fusion) two or more vertebrae are immobilized by inserting pedicle screws in the vertebrae and linking the pedicle screws using metal wire. The technique is often used to treat degenerative diseases like scoliosis, degenerative disc disease, spinal stenosis, and tumors. The use of spondylodesis in the US has increased over the last two decades and is accelerating [Deyo et al., 2005, 2004] with the US having the highest incidence rate [Deyo & Mirza, 2006]. In Denmark 1386 operations were performed in 2005 increasing to 1628 operations in 2006 [Rasmussen et al., 2009].

In the conventional surgical technique per-operative spinal navigation is based on both the surgeon's anatomical knowledge and pre-operative images. However, this navigation is difficult causing potentially severe errors in screw placement. Using manual navigation, error rates in the range of 20–30% have been reported [Koller et al., 2008; Neo et al., 2005; Rajasekaran et al., 2007]. The error rate depends heavily on the surgeon's level of experience. Very experienced surgeons can achieve significantly lower error rates [Kotil & Bilge, 2008]. Using image guiding or computer assistance the error rate is also reduced [Fu et al., 2004; Ito et al., 2008; Kotani et al., 2007; Laine et al., 2000; Ludwig et al., 2000].

In computer-assisted spondylodesis the surgical tools are tracked in real-time visualizing their location and orientation in pre-operative CT images. This requires registration of the pre-operative images to the patient during surgery. Currently this is done manually and takes approximately 10 min per vertebra [Arand et al., 2002], a significant part of the procedure. Because the spinal column is flexible the vertebrae move with respect to each other. Per-operative freehand ultrasound (US) can be used to verify the configuration of the verte-

brae during surgery. An automatic technique for US-CT registration has been developed [Mercier et al., 2005, 2011; Yan et al., 2011]. The US-CT registration technique requires segmentation of the individual vertebrae from CT images, a task that is complicated by the fact that CT images show no clear boundary between the vertebrae.

3.2 Computed tomography of the vertebrae

Examples of the appearance of a vertebra in CT images is shown in Figure 3.2. The dense wall of the vertebra has a CT value of above 200 HU, while the interior body consisting of cancellous bone has a value of 0–150 HU.

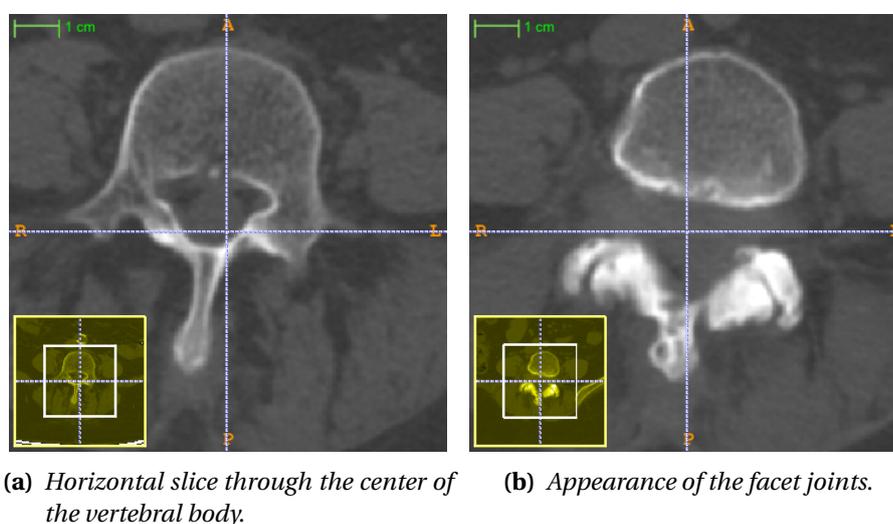


Figure 3.2: CT appearance of L4. Images from the data set.

The intervertebral disk separating adjacent vertebral bodies has CT intensity values in the range 50–100 HU, making it easy to segment from the denser wall of the vertebral body. The facet joints appear as thin darker fissures between the dense articulate processes, but the fissures are so narrow that they can partly disappear on CT images due to the partial volume effect (Fig. 3.2(b)).

3.3 Prior work in vertebra segmentation

Prior work in vertebral segmentation has focused mainly on applying elastic models to X-ray or CT images, reporting segmentation errors of around 1 mm [Cohen et al., 1992; Klinder et al., 2008; Mastmeyer et al., 2006; Shen et al., 2008; Weese et al., 2001], or building statistical shape models from CT data [Becker et al., 2011; Kaus et al., 2003; Kirschner et al., 2011; Klinder et al., 2009; Lorenz & Krahnstover, 1999].

Mastmeyer et al. [2006] applied a 3-D deformable model for segmentation of the vertebral body, achieving a Dice kappa of 0.99 on phantom data. Kirschner et al. [2011] evaluated different implementations of active shape models on thoracic and lumbar vertebrae, achieving a Hausdorff distance of 6.88 ± 2.45 mm

Other authors have combined morphology, region growing and other methods into hybrid methods for vertebral body segmentation [Aslan, Ali, Chen, et al., 2010; Aslan et al., 2011; Aslan, Ali, Farag, et al., 2010] or segmentation of full vertebrae [Kim & Kim, 2009]. A Jaccard similarity coefficient of 0.95 for vertebral body segmentation has been reported [Aslan, Ali, Chen, et al., 2010]. Also some work has been done in segmenting X-ray images using either 2-D [Roberts et al., 2009] or 3-D shape models [Benameur et al., 2003].

So far no work has been published on applying active appearance models to the problem of 3-D segmentation of individual vertebrae from CT images of the spine.

PART



Methods

Data and tools

The segmentation method was evaluated using prostate MRI data provided by the Department of Medical Physics at Aalborg Hospital, Denmark, and vertebra CT images provided by McConnell Brain Imaging Center, Montreal Neurological Institute, McGill University, Canada.

Prostate images

The prostate images were axial T2 weighted fast relaxation fast spin echo (FRFSE) MRI data from 25 patients. The images were acquired as part of radiotherapy dose planning for treatment of prostate cancer. As part of the clinical procedure a stent had been inserted in the prostatic urethra for automatic MRI-CT registration [Østergaard et al., 2010]. The stent is not used in the current study.

The images contain manual delineations of the prostate and other anatomical structures relevant for radiotherapy planning, but only the prostate delineations are used in this study. The images were provided as anonymized DICOM format slices with delineations represented as DICOM-RT structures. The delineations are a set of slice-by-slice contours defined by a series of points in the world coordinate system. Thus the manual segmentations have sub-voxel precision, however not necessarily sub-voxel accuracy. Each of the 25 patients have had two MRI scans conducted as part of the radiotherapy, thus a total of 50 image volumes with delineations are provided.

Each image slice is 512×512 voxels with a voxel size of 0.5469 to 0.5664 mm. The volumes range from 16 to 40 slices with a slice thickness of 3 to 3.3 mm.

The manual segmentations had been made for radiotherapy planning purposes and were not originally intended for use as the basis of a statistical model. The manual segmentations are the clinical target volume, which does not necessarily contain only the prostate or even the full prostate. The scans from four patients were excluded because the delineations clearly contained other anatomical structures or pathologies in addition to the prostate:

- Segmentation for patient 11 includes extra-prostatic tissue.
- Segmentation for patient 19 includes the seminal vesicles.
- Segmentation for patient 23 includes extra-prostatic tissue.

- Segmentation for patient 24 includes part of one seminal vesicle.

These patients were excluded because the statistical model should only describe the variability of the prostate, and the extra tissue included in the manual segmentations would otherwise affect the statistical model and method validation. After exclusion a total of 42 prostate image volumes were used in this study.

Vertebra images

CT images of the L4 vertebra from 27 patients were provided; one image volume for each patient. Images were acquired with 130 kVp tube potential and 175 A tube current. One image was acquired with 225 A tube current. Patients are 22 females and 5 males, 48–79 years old (median 67).

A manual voxel-based segmentation of the L4 vertebra was provided for each image. The segmentations were made for this particular study. Because the manual segmentation is voxel-based the precision is limited by the voxel size. The in-plane resolution is 512×512 voxels with a voxel size of 0.352×0.352 mm. The number of slices range from 55 to 200 with a slice thickness of 0.998 to 2.002 mm. The images and segmentations were provided in MINC format.

4.1 Programming framework

The appearance model building and iterative segmentation algorithm was implemented in MATLAB.

Image preprocessing and segmentation performance evaluation was done using the NIAK image analysis kit [Bellec, 2011a] and PSOM pipelining system [Bellec, 2011b; Bellec et al., 2012] for MATLAB, as well as the MINC imaging tools [McConnell Brain Imaging Center, 2012]. The following MINC packages were used:

MINC: Main package and DICOM to MINC conversion.

EZminc: Resampling, distance map computation and metrics.

conglomerate: MNI .obj format¹ to MINC conversion and vice versa.

bicpl: MNI .obj format tools.

mni_autoreg: Automatic registration and volume cropping.

¹The MNI .obj format was developed by David MacDonald at Montreal Neurological Institute, and is used for representing objects within the MINC toolkit. Objects are represented by vertex coordinates, faces, and point normals. A reference description of the file format is available at http://www.bic.mni.mcgill.ca/users/mishkin/mni_obj_format.pdf.

Appearance modeling

A model of a class of objects has to be based on a certain way of representing the object. One approach is to represent the object by its geometrical properties, i.e. shape and size. Using this approach the shape description has to be extracted from images of the object. Depending on the type of model and shape description this can be done automatically or manually for a set of training data.

Another approach is to represent the object directly from the texture in images of the object, eliminating the need for extraction of shape descriptors. Models built using this approach do, however, not contain any description of shape, and can thus not be used for segmentation.

In the Active Appearance Model (AAM) a shape and a texture model of the class of objects are combined. The models are statistical models based on Principal Component Analysis (PCA) [Lay, 2006], modeling the possible modes of variation from the average object. That is; the shape model contains the mean shape, the possible modes of variation seen in the training data set (eigenvectors), and the typical range of these modes of variation (eigenvalues), and likewise for the texture model. The idea behind the Active Appearance Model is the reasonable assumption that if the shape of an object changes, so does the grayscale intensities in the image of the object [Cootes et al., 1998]. Thus the shape and texture models are combined using PCA, resulting in a statistical model of appearance describing both object shape and texture with the same parameters.

The Active Appearance Model is a generative model; by varying the parameters of the AAM new images of objects from the class of objects in the training data set can be generated. At the same time a representation of the shape of the object in the new image is generated by the model. In this way the model can be used for segmenting previously unseen images by finding the model parameters that generate an image with the best possible match to the unseen image, and then extracting the shape from the model.

5.1 Shape representation

The traditional approach to representing object shape is to identify a set of landmark points on the surface of the object. A landmark can be defined as an anatomical landmark, a mathematical land-

mark or a pseudo-landmark. The placement of anatomical landmarks is defined by the anatomy of the object and located at anatomically meaningful points. Mathematical landmarks are defined by geometrical properties such as high curvature or extreme points. Pseudo-landmarks are constructed between anatomical or mathematical landmarks and can be used for approximating continuous curves. Landmarks for shape representation must be labeled so that corresponding landmarks in different images share the same label. [Dryden & Mardia, 1998]

From a set of labeled training images a point distribution model (PDM) for the shape of the class of objects can be built [Cootes et al., 1992]. The model describes the mean landmark location and variability by applying PCA on the landmark coordinates. It is important that the training images are aligned and that landmarks correspond to the same anatomical position in each training image. Establishing the correspondence manually is a tedious task, especially for 3-D images, and automatic methods are prone to errors. Furthermore, point distribution models are not very good at handling topological changes and are prone to numerical instability. [Stegmann & Gomez, 2002]

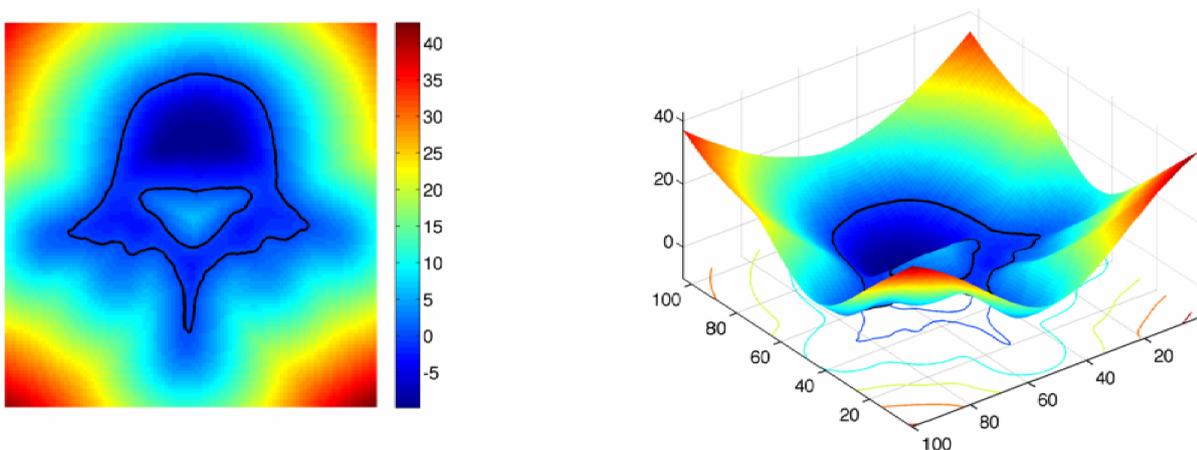


Figure 5.1: Example distance map of a vertebra. Left: A single slice of the distance map with the zero level set outlined as a black contour. Right: The same slice represented as a 3-D surface.

A more recent approach to shape representation is to represent the shape as the zero level set of a signed distance map [Leventon et al., 2000; Tsai et al., 2003]. The signed distance map is the shortest distance from any point in the image to the surface of the object, with outside points having positive values and inside points negative (Fig. 5.1) [Borgefors, 1991]. Points having a value of zero thus implicitly represent the surface of the object. A shape model based on this shape representation also need correspondence between train-

ing images, but this can be achieved by aligning the training images before building the model. The alignment will not produce perfect correspondence between the training images as the point-wise matching of landmarks will. However, the level set shape representation is quite robust to small misalignments, as neighboring voxels in the distance map are highly correlated because the distance only varies slowly across the image volume [Leventon et al., 2000].

5.2 Texture

A model of image texture can be built using the grayscale values directly, as is done in the eigenface approach [Kirby & Sirovich, 1990; Murase & Nayar, 1995; Turk & Pentland, 1991]. It can also be built using mathematically derived texture features like gradient images or other texture enhancing operations. An example of this is the use of Haar wavelets to extract texture features of interest [Ghose et al., 2010].

Like with the shape model it is important that texture features are comparable across training images [Cootes & Taylor, 2004]. Thus when building a model using grayscale intensity values directly as a texture feature the training images must be normalized to remove global illumination differences and shape. In this way the training images are transformed into the same grayscale intensity and shape reference frame.

In MRI the grayscale intensity values are computed using the inverse Fourier transform, thus the mean is completely arbitrary. The grayscale intensity variance depends on magnetic and radiofrequency field strength, and thus does not provide any meaningful information for segmentation. For MR images the grayscale intensity is normalized to zero mean and a set standard deviation. The standard deviation can be chosen arbitrarily. In this project the standard deviation was set at 450, which is similar to the standard deviation in the training images. Each training image \vec{I} is normalized by applying a scaling and offset:

$$\vec{g} = (\vec{I} - \hat{I}) \frac{450}{\sigma_I} \quad (5.1)$$

The offset \hat{I} is the mean grayscale intensity, σ_I is the grayscale intensity standard deviation, and \vec{g} is the grayscale normalized texture. For CT images this grayscale normalization is not necessary as the grayscale intensities are given in standardized values (Hounsfield Units).

Shape normalization is done by aligning the training images in the same way as when building the distance map shape model described above.

5.3 PCA-based models

A statistical model of an image feature is built from a vector of feature values \vec{x}_i ; one vector for each image i in the training data set of n images. Depending on the image feature to model this vector can be a vector of landmark coordinates, voxel grayscale intensity values, voxel distance map values, etc. For each training image the feature offset $\vec{\delta x}_i$ is calculated by subtracting the mean feature vector $\hat{\vec{x}}$:

$$\hat{\vec{x}} = \frac{1}{n} \sum_{i=1..n} \vec{x}_i \quad (5.2)$$

$$\vec{\delta x}_i = \vec{x}_i - \hat{\vec{x}} \quad (5.3)$$

The set of all feature offset vectors from the training data set constitute the variability matrix $\mathbf{X} = [\vec{\delta x}_1 \cdots \vec{\delta x}_n]$ of size $m \times n$, where m equals the number of feature elements (i.e. landmark coordinates or image voxels) and n is the number of training image volumes.

Using Singular Value Decomposition (SVD) [Lay, 2006] the matrix \mathbf{P}_x of eigenvectors is computed:

$$\mathbf{X}\mathbf{X}^\top = \mathbf{P}_x \Sigma_x \Sigma_x^\top \mathbf{P}_x^\top \quad (5.4)$$

\mathbf{P}_x is the eigenvectors of $\mathbf{X}\mathbf{X}^\top$. The diagonal of Σ_x contains the square root of the corresponding eigenvalues. The eigenvectors are the principal modes of variation of the image feature, and the eigenvalues express the typical range of variation of each mode in the training data set.

A new feature vector can be generated from the eigenvectors by multiplying a parameter vector \vec{b}_x and adding the mean feature vector $\hat{\vec{x}}$:

$$\vec{x} = \hat{\vec{x}} + \mathbf{P}_x \vec{b}_x \quad (5.5)$$

By constraining the parameter vector \vec{b}_x to within ± 3 standard deviations of each parameter (the diagonal of Σ_x) only plausible feature vectors will be generated. [Cootes et al., 1998]

Spurious variations and noise in the training data set can be filtered by removing the eigenvectors with the smallest corresponding eigenvalues, retaining only a certain fraction of the total variance in the model. This also reduces the memory footprint of the model.

5.4 Level set appearance model

Utilizing the shape, texture, and statistical modeling concepts described above a full appearance model is built. The appearance model implemented in this project is based on a level set description of shape and a direct grayscale value representation of texture [Hu & Collins, 2007].

The shape model is generated from the set of distance maps $\vec{\phi}_i$ in the training data set, resulting in the principal modes of shape variation \mathbf{P}_s . Likewise a texture model is generated from the set of grayscale intensity values \vec{g}_i , resulting in the principal modes of texture variation \mathbf{P}_g . The full appearance (shape and texture) of an unseen image can be described by the shape and texture parameters \vec{b}_s and \vec{b}_g . In the appearance model these parameters are combined into the appearance parameters \vec{c} .

A parameter matrix \mathbf{B} is determined from the shape and texture parameters describing the training data set:

$$\mathbf{B} = \begin{bmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_g \end{bmatrix} = \begin{bmatrix} \mathbf{W}_s \mathbf{P}_s^\top \mathbf{S} \\ \mathbf{P}_g^\top \mathbf{G} \end{bmatrix} \quad (5.6)$$

The shape variability matrix \mathbf{S} and texture variability matrix \mathbf{G} are computed from the training data set.

The weight matrix \mathbf{W}_s is multiplied to account for the difference in units between shape and texture. This weighing makes the shape and texture parameters approximately the same scale. It is defined by the ratio of the total grayscale intensity variance $\Sigma_g \Sigma_g^\top$ to the total shape variance $\Sigma_s \Sigma_s^\top$ ($\mathbb{1}$ is the identity matrix):

$$\mathbf{W}_s = \mathbb{1} \sqrt{\frac{\Sigma(\Sigma_g \Sigma_g^\top)}{\Sigma(\Sigma_s \Sigma_s^\top)}} \quad (5.7)$$

The principal modes of appearance variation \mathbf{P}_c can be computed by applying PCA on the parameter matrix \mathbf{B} containing the shape and texture parameters of the training data set.

$$\begin{aligned} \mathbf{B} &= \mathbf{P}_c \mathbf{c} \\ \begin{bmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_g \end{bmatrix} &= \begin{bmatrix} \mathbf{P}_{cs} \\ \mathbf{P}_{cg} \end{bmatrix} \mathbf{c} \end{aligned} \quad (5.8)$$

Combining (5.5) and (5.8), new images and shapes can be synthesized using the appearance model by varying the appearance parameters \vec{c} :

$$\begin{bmatrix} \vec{\phi} \\ \vec{g} \end{bmatrix} = \begin{bmatrix} \hat{\vec{\phi}} \\ \hat{\vec{g}} \end{bmatrix} + \begin{bmatrix} \mathbf{Q}_s \\ \mathbf{Q}_g \end{bmatrix} \vec{c} \quad (5.9)$$

The variability matrices \mathbf{Q}_s and \mathbf{Q}_g are:

$$\begin{bmatrix} \mathbf{Q}_s \\ \mathbf{Q}_g \end{bmatrix} = \begin{bmatrix} \mathbf{P}_s \mathbf{W}_s^{-1} \mathbf{P}_{cs} \\ \mathbf{P}_g \mathbf{P}_{cg} \end{bmatrix} \quad (5.10)$$

In summary the appearance model consists of the mean shape and texture ($\hat{\vec{\phi}}$ and $\hat{\vec{g}}$) plus the modes of shape and texture variation (\mathbf{Q}_s and \mathbf{Q}_g). Plausible images and shapes can be generated by constraining values of \vec{c} to the interval between -3 and +3 standard deviations. [Cootes et al., 1998]

Segmentation

The goal when using the appearance model for image segmentation is to find the model parameters which generate a model image as close to the target image as possible. This matching of the model to a target image can thus be treated as an optimization problem in which the squared difference $|\Delta\vec{I}|^2$ between the target image \vec{I}_s and the image generated by the model \vec{I}_m is minimized:

$$\Delta\vec{I} = \vec{I}_s - \vec{I}_m \quad (6.1)$$

Because the appearance model typically has many parameters this appears to be a high dimensional optimization problem. However, each attempt at solving the problem is a similar optimization problem. Cootes et al. [1998] point out that the spatial pattern of $\Delta\vec{I}$ encodes information about how to adjust the model parameters, and suggest to use this information to learn in advance how to adjust the parameters in order to achieve a better model fit.

The target and model images are not necessarily comparable in grayscale intensity values and 3-D pose. Thus the images must be transformed into a common grayscale intensity reference frame in which the mean and variance of the grayscale intensity values are comparable. Likewise the images must be transformed into a common coordinate system to achieve the same 3-D pose.

The problem of applying the appearance model for image segmentation can be broken down into the following problems:

- Establish transformations of grayscale intensity and 3-D pose between the target and model image reference frame.
- Learn the relationship between $\Delta\vec{I}$ and a change in model parameters in advance.
- Use this a priori information for image segmentation.

Each of these problems are addressed in turn in the following.

6.1 Transformation between reference frames

Transformation between image and model coordinate systems is done through rotation, scaling and translation. To transform the target image into the model coordinate system the target image is sampled at points defined by the transformation matrix $\mathbf{T}(\vec{t})$, where \vec{t} is

the pose parameter vector:

$$\vec{t} = [\phi, \theta, \psi, s, t_x, t_y, t_z]^T \quad (6.2)$$

The transformation is defined so that the zero vector represents no change in pose ($\mathbf{T}(\vec{0}) = \mathbb{1}$), and so that for small pose changes $\mathbf{T}(\vec{t}_1)\mathbf{T}(\vec{t}_2) \approx \mathbf{T}(\vec{t}_1 + \vec{t}_2)$. These properties are necessary for the prediction of pose parameter adjustments to be consistent [Cootes & Taylor, 2004]. The properties are straightforward to achieve for rotations and translations. In order to achieve these properties for scale the parameter s is chosen to be the natural logarithm of the actual scaling. For details of the pose representation see Appendix A.

To transform the target image grayscale intensity values into the model grayscale intensity reference frame the target image is normalized to zero mean, and a scaling is applied on the grayscale intensity values:

$$\vec{g}_s = (\vec{I}_s - \hat{I}_s) \frac{1}{1 + u} \quad (6.3)$$

The texture scale parameter u is defined so that $u = 0$ represents no change in grayscale intensity to achieve the same properties of the transformation as described above.

Combining the appearance parameters \vec{c} , the pose parameters \vec{t} , and the texture parameter u , the full set of model parameters \vec{p} are:

$$\vec{p} = \begin{pmatrix} \vec{c} \\ \vec{t} \\ u \end{pmatrix} \quad (6.4)$$

6.2 Learning parameter adjustments

Using the full model parameters \vec{p} the optimization problem is reformulated to minimize the square error $E(\vec{p}) = |\vec{r}(\vec{p})|^2$ of the residuals \vec{r} , where:

$$\vec{r}(\vec{p}) = \vec{g}_s(\vec{t}, u) - \vec{g}_m(\vec{c}) \quad (6.5)$$

$$\vec{g}_m(\vec{c}) = \hat{\vec{g}} + \mathbf{Q}_g \vec{c} \quad (6.6)$$

The synthesized model image \vec{g}_m is defined in (5.9) which is repeated above, and the sampled target image \vec{g}_s is the target image sampled at points defined by the 3-D pose parameters \vec{t} and transformed into the model grayscale intensity reference frame by (6.3).

A first order Taylor expansion of $\vec{r}(\vec{p})$ around \vec{p} gives:

$$\vec{r}(\vec{p} + \Delta\vec{p}) = \vec{r}(\vec{p}) + \frac{\partial \vec{r}}{\partial \vec{p}}(\vec{p}) \Delta\vec{p} \quad (6.7)$$

The (i, j) th element of the matrix $\frac{\partial \vec{r}}{\partial \vec{p}}$ is the partial derivative $\frac{\partial r_i}{\partial p_j}$.

To achieve a better model fit in an iterative optimization framework the objective is to find the parameter displacements $\Delta\vec{p}$ which minimize $|\vec{r}(\vec{p} + \Delta\vec{p})|^2$, where \vec{p} is the initial parameter estimates. Setting (6.7) equal to zero the linear least squares solution gives:

$$\Delta\vec{p} = -\mathbf{R}\vec{r} \quad (6.8)$$

The parameter change prediction matrix \mathbf{R} is the left inverse of the partial derivative matrix $\frac{\partial\vec{r}}{\partial\vec{p}}$:

$$\mathbf{R} = \left(\frac{\partial\vec{r}^\top}{\partial\vec{p}} \frac{\partial\vec{r}}{\partial\vec{p}} \right)^{-1} \frac{\partial\vec{r}^\top}{\partial\vec{p}} \quad (6.9)$$

The matrix \mathbf{R} depends on the current parameter estimates \vec{p} and should normally be recalculated at every step. However, \mathbf{R} is assumed to be relatively fixed, and can be estimated from the training data set by displacing each parameter in turn from the known optimal value and average over the training set [Cootes & Taylor, 2004].

The partial derivative for each parameter is estimated by a Gaussian weighted sum of derivatives for different displacements around zero:

$$\frac{\partial r_i}{\partial p_j} = \sum_k w(k) \frac{r_i(\vec{p} + \delta p_{jk}) - r_i(\vec{p})}{\delta p_{jk}} \quad (6.10)$$

The weighting factor $w(k)$ is a Gaussian window normalized to make $\sum_k w(k) = 1$, where k is the index of the current displacement. By applying the Gaussian window the derivatives for small displacements are weighted higher than larger displacements in the derivative estimation.

For the training data set the optimal values of the pose and texture parameters are the zero vector, because the training volumes are already normalized in pose and grayscale texture. Thus the calculation of the residuals displacement for pose parameters reduce to:

$$\begin{aligned} \Delta\vec{r}_{\vec{t}} &= \vec{r}(\vec{p} + \Delta\vec{t}) - \vec{r}(\vec{p}) \\ &= \vec{g}_s(\Delta\vec{t}, 0) - \vec{g}_m(\vec{c}) - \vec{g}_s(\vec{0}) + \vec{g}_m(\vec{c}) \\ &= \vec{g}_s(\Delta\vec{t}, 0) - \vec{g}_s(\vec{0}) \end{aligned} \quad (6.11)$$

A similar result is achieved for the texture parameter u . The derivative for the pose and texture parameters can thus be estimated as an average over the training data set, without even knowing the appearance model. This implies that the part of $\frac{\partial r_i}{\partial p_j}$ corresponding to the pose and texture parameters can be computed once for each training image, and does not need to be recomputed if the choice of parameters for model building change.

The calculation of the residuals displacement for the appearance parameters \vec{c} reduce to:

$$\begin{aligned}
\Delta \vec{r}_{\vec{c}} &= \vec{r}(\vec{p} + \Delta \vec{p}) - \vec{r}(\vec{p}) \\
&= \left(\vec{g}_s - \widehat{\vec{g}} - \mathbf{Q}_g(\vec{c} + \Delta \vec{c}) \right) - \left(\vec{g}_s - \widehat{\vec{g}} - \mathbf{Q}_g \vec{c} \right) \\
&= \mathbf{Q}_g \vec{c} - \mathbf{Q}_g(\vec{c} + \Delta \vec{c}) \\
&= -\mathbf{Q}_g \Delta \vec{c}
\end{aligned} \tag{6.12}$$

This can be estimated once directly from the model, and does not need to be computed for each image in the training data set.

The range of parameter displacements for estimating the derivative $\frac{\partial r_i}{\partial p_j}$ is chosen based on the suggestions by Cootes & Taylor [2004]. The chosen ranges are:

Appearance \vec{c} : ± 0.5 times the standard deviation (square root of the eigenvalue) of each parameter.

Rotations ϕ, θ, ψ : $\pm \pi$, i.e. the full range of rotations.

Scale s : ± 0.1 , corresponding to the range 0.905–1.105 (or about $\pm 10\%$) in actual scaling.

Translations t_x, t_y, t_z : ± 3 voxels, corresponding to ± 3 mm.

Texture scale u : ± 0.1 , corresponding to $\pm 10\%$ in texture scaling.

6.3 Iterative segmentation algorithm

The learned relationship \mathbf{R} between the residuals \vec{r} and the change in parameters $\Delta \vec{p}$ is used in an iterative algorithm for matching the model to a target image (Algorithm 6.1). The algorithm is comparable to gradient descent; the difference being that the gradient direction is estimated using prior knowledge instead of being calculated in every step.

The algorithm takes the target image \vec{I}_s , the appearance model, the parameter change prediction matrix \mathbf{R} , and an initial parameter guess \vec{p} as inputs.

The initial guess of parameters could be the zero vector for appearance parameters \vec{c}_0 and texture scale u_0 . The pose parameters \vec{t}_0 should be initialized to a qualified guess of the location and pose of the object of interest in the target image. One choice could be zero change in rotations and scale and to initialize translation to the center of the target image. The translations could also be a user-supplied manual estimate of the location of the modeled object in the target image.

The algorithm is initialized by computing the residuals \vec{r} and squared error E for the initial parameter estimates (line 1–2). For each iteration the proposed change in parameters $\Delta \vec{p}$ is computed (line 4). The

Algorithm 6.1 Iterative segmentation algorithm using predictions of parameter changes. Input is the target image \vec{I}_s , the appearance model $\hat{\vec{g}}$ and \mathbf{Q}_g , the parameter change prediction matrix \mathbf{R} , and an initial parameter estimate \vec{p} , which could just be the zero vector. The algorithm will change the parameters in the direction suggested by \mathbf{R} for different step sizes, but only as long as the parameter change improves the overall grayscale fit between the model and target image.

Require: $\vec{I}_s, \hat{\vec{g}}, \mathbf{Q}_g, \mathbf{R}, \vec{p} = (\vec{c}_0, \vec{t}_0, u_0)^\top$

```

1:  $\vec{r} \leftarrow \hat{\vec{g}}_s(\vec{t}_0, u_0) - \hat{\vec{g}}_m(\vec{c}_0)$ 
2:  $E \leftarrow \vec{r}^\top \vec{r}$ 
3: repeat
4:    $\Delta\vec{p} \leftarrow -\mathbf{R}\vec{r}$ 
5:   for  $k = [1, 1.5, 0.5, 0.25, \dots]$  do
6:      $\vec{p}' \leftarrow \vec{p} + k\Delta\vec{p}$ 
7:     Constrain  $\vec{p}'$  to plausible values.
8:      $\vec{r}' \leftarrow \hat{\vec{g}}_s(\vec{t}', u') - \hat{\vec{g}}_m(\vec{c}')$ 
9:      $E' \leftarrow \vec{r}'^\top \vec{r}'$ 
10:    if  $E' < E$  then
11:       $E \leftarrow E'$ 
12:       $\vec{p} \leftarrow \vec{p}'$ 
13:      break for-loop
14:    end if
15:  end for
16: until  $E' > E$ 
17: return  $\vec{p}$ 

```

proposed change $\Delta\vec{p}$ is the direction in the high dimensional parameter space in which the squared error is reduced.

A simplistic line search is performed to allow for a variable step size k (line 5). Different pre-defined step sizes are tried in turn until one is found which reduces the squared error E . For each step size the new parameter values are computed from the estimated adjustments (line 6), and the appearance parameters are constrained to within ± 3 standard deviations of each appearance parameter. This is done to ensure that only plausible shapes will be generated.

The updated squared error E' is computed. If E' is smaller than the previous squared error E the estimate is accepted as an improvement, and the line search is terminated to move to the next iteration step. If the line search go through all step sizes without finding any improvement of the squared error then convergence is declared and the algorithm is terminated.

Image pre-processing

To build an appearance model the training images have to have the same pose and be comparable in grayscale intensity characteristics. The goal of image pre-processing is to register all training images to a common volume of reference in order to remove pose difference among training sets, to normalize the grayscale intensity, and to convert images to a common format for further analysis.

Training image volumes for model building must be:

- Comparable in grayscale intensity characteristics.
- Registered in a common coordinate system.
- Represented in the same file format.

Manually segmented training shapes for model building must be:

- In the same common coordinate system as the images.
- Represented as a signed distance map.

The exact pre-processing steps necessary depend on the imaging modality and file format of the original images.

7.1 Common reference

The training volumes are registered to a common reference volume (Fig. 7.1(a)). This reference is manually created based on one of the training images.

To create the common reference the image is rotated to make the transverse plane of the patient correspond to the x-y plane in the image and to make the sagittal plane correspond to the y-z plane. The rotation center is set to the center of the manual segmentation (mean of contour point coordinates), and the rotation parameters are determined manually by inspecting the image.

The image volume and segmentation shape are transformed according to the chosen rotation.

7.2 Registration

Training images are registered to the common reference to remove differences in pose. Registration was done using the minctracc tool

[Collins et al., 1994]. The registration is done in three steps presented below:

1. Remove translations using the location of the manual segmentation.
2. Roughly remove rotations using the full image information.
3. Fine-tune rotations and scale using only image information around the segmentation.

The first step is a simple computation, and only the last two steps use `minctracc`. These last two registration steps are based on Mutual Information [Collignon et al., 1995], and the center of rotation is set to the center of the manual segmentation. An example of the result of each step is shown in Figure 7.1 and 7.2, along with the common reference.

The translation is removed by computing the transformation required to move the segmentation center to the center of the reference segmentation. This makes a near-perfect alignment in translation, as it is based on the manual segmentation. This transformation is used as the initial transformation guess in the next step.

In step two a six parameter rigid body transformation (3 translations, 3 rotations) is computed using a simplex radius of 5 on a 3-D lattice with a step size of 4 mm. Because this initial transformation guess is quite good, translations are weighted by a factor of 0.01 to punish large translations away from the initial guess. Thus the main function of this step is to align the images in terms of rotations (Fig. 7.1(c)).

In the final registration step both the reference and the image volume are masked to a volume of interest (VOI) defined by the corresponding manual segmentation dilated by a 26-connected kernel of $40 \times 40 \times 40$ mm. This creates a VOI containing the segmented region plus a margin of 2 cm. The actual dilation kernel dimensions depend on the voxel dimensions. The transformation result from the previous step is used as the initial guess. A seven parameter procrustes transformation (3 translations, 3 rotations, 1 scale) is computed using a simplex radius of 1 on a 3-D lattice with the step size reduced to 2 mm. Again translations are weighted by 0.01, and scaling is weighted by 0.005 making a 0.5% change in scale weighted equal to 1 degree of rotation. This final step will fine-tune the rotations while allowing for some scaling (Fig. 7.2(d) and 7.1(d)).

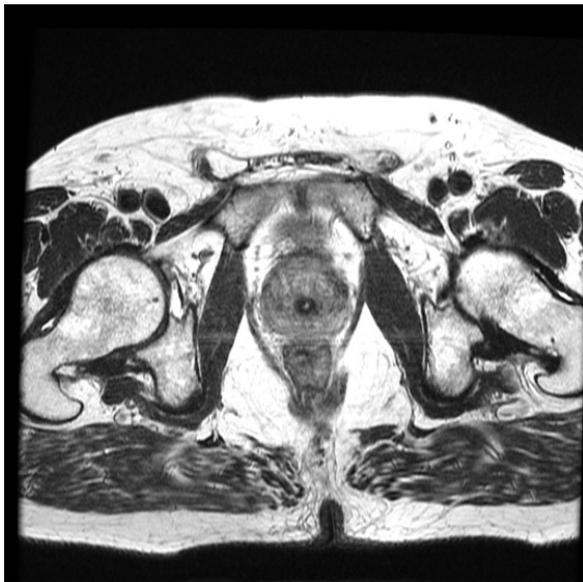
The result of the registration is a transformation matrix which can be used for transforming the image volume and segmentation into the common space of the reference.



(a) Common reference.



(b) After the first step; translations removed, initial rotations kept.



(c) After the second step; rotations removed.



(d) After the third step; rotations and scale further refined.

Figure 7.1: Example of image registration after each step (transverse plane). The top left image is the common reference. Each image is in the space of the common reference. Notice how the image is rotated first to roughly align the overall image (c), and later to finely align the prostate and surrounding tissue (d), decreasing the overall match but improving the local match around the prostate. Image is prostate cancer patient 12 scan 1.

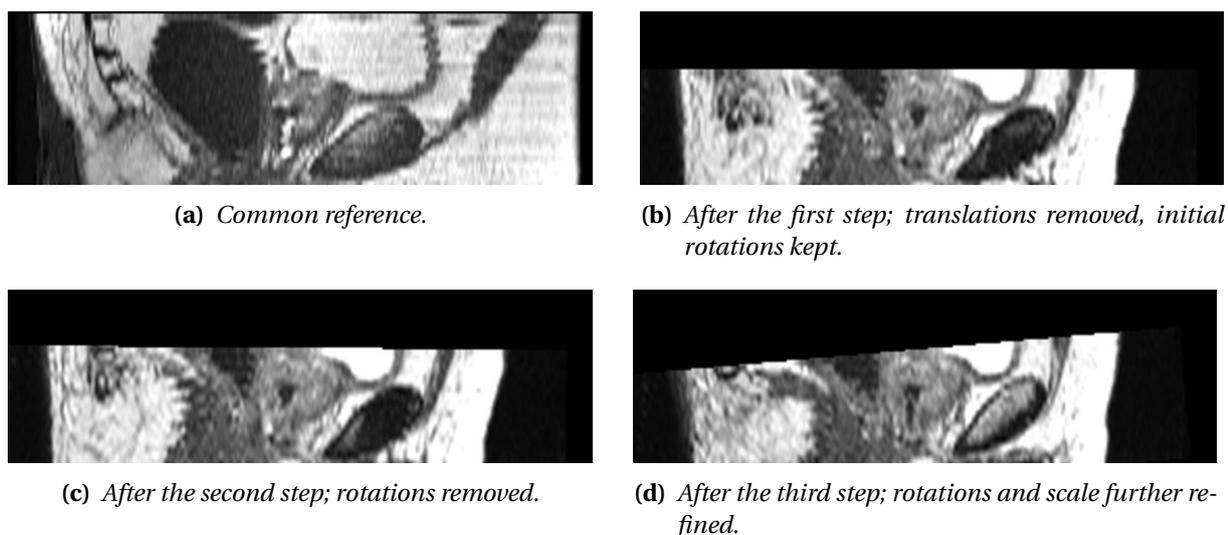


Figure 7.2: Example of image registration after each step (sagittal plane). The top left image is the common reference. Each image is in the coordinate system of the common reference. Notice how the image is translated (b) to align with the reference (a). Image is prostate cancer patient 12 scan 1.

7.3 Image volumes

Image volumes are preprocessed in the following steps:

1. Convert image volumes to the MINC file format.
2. (*MRI only*) Normalize grayscale intensities to zero mean and a defined variance.
3. Transform the image volume into the common space and re-sample to isotropic resolution.
4. Crop the volume to a VOI defined from manual segmentations.

First the image volumes are converted to the MINC file format.

For MR images the grayscale intensity values have to be normalized. This is not necessary for CT images as the grayscale intensities are given in Hounsfield Units. Details of normalization are given in section 5.2.

The image volume is then transformed into the space of the common reference using the transformation computed during registration (section 7.2). The image is resampled to an isotropic resolution of 1 mm using B-spline interpolation.

Finally the image is cropped to a volume of interest defined by the union of all training segmentations.

7.4 Manual segmentations

Manual segmentations represented as DICOM-RT structure sets are pre-processed in the following steps:

1. Convert DICOM-RT structures to an MNI .obj format object.
2. Transform the object into the common space.
3. Convert the object to a high-resolution binary MINC mask cropped to the desired VOI.
4. Compute the signed distance map from the high-resolution binary volume.
5. Downsample the distance map to normal resolution.

Examples of the original DICOM-RT shape representation, the MNI .obj representation (step 1), and the distance map representation (step 4–5) is shown in Figure 7.3.

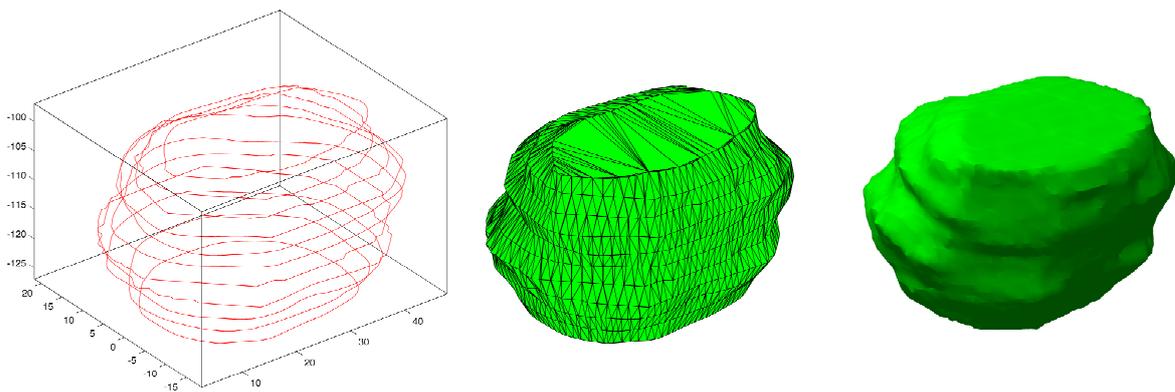


Figure 7.3: Example of a manual segmentation represented as slice contours, a triangular surface mesh (MNI .obj file format), and the zero level set of a signed distance map, respectively. Based on prostate cancer patient 4 scan 1.

In DICOM-RT the segmentations are represented as a series of contour points defined slice-by-slice (Fig. 7.3). The points are extracted from the DICOM-RT header and converted from the DICOM left-posterior-superior coordinate system to the MINC right-anterior-superior coordinate system. A Delaunay triangulation of the points is computed, and the resulting tetrahedra are filtered to remove tetrahedra lying outside of the segmentation contours. For each tetrahedron the circumcenter is computed and projected onto the nearest x-y plane (image slice). If the circumcenter is outside of the segmentation contour of the slice the tetrahedron is discarded. Next a triangular surface mesh of the remaining tetrahedra is computed (Fig. 7.3). The MNI .obj file format contains the mesh points, faces and point

normals. The point normals are determined as a weighted average of the face normals weighted by the reciprocal euclidean distance between the mesh point and the face circumcenter. The mesh points, faces and point normals are written in the MNI .obj file format.

The object is transformed into the space of the common reference using the transformation computed during registration (section 7.2).

The transformed object is converted to a high-resolution binary MINC mask based on a cropped version of the reference volume with an isotropic resolution of 0.2 mm. The resolution is a compromise between the desired precision of the final distance map and the available amount of RAM for computation. A higher resolution will reduce the uncertainties introduced by the voxelation of the shape. Cropping is done also to reduce the memory requirement. The cropped region is the same as for the image volumes (section 7.3).

The signed distance map is computed from the binary mask. It is defined as the distance from the current voxel to the surface of the segmentation with outside voxels having positive values and inside voxels negative. Thus the surface of the segmented object is represented by the zero level-set of the signed distance map (section 5.1 and Figure 7.3).

Finally the high-resolution distance map is downsampled to the resolution of the final normalized and cropped image volume (section 7.3).

If the segmentations are provided as binary volumes the distance map is computed directly from the binary volumes and cropped to the desired VOI.

7.5 Preprocessing pipeline implementation

The registration, image preprocessing, and manual segmentation preprocessing steps are dependent upon each other; registration cannot be done without a MINC image and an object representation of the manual segmentation, and the image and segmentation preprocessing cannot complete without the registration. Also the VOI for cropping is determined from the manual segmentations.

To handle all the steps and dependencies the preprocessing tasks are built into a PSOM pipeline. For an overview of the DICOM preprocessing pipeline see Figure 7.4. The pipeline will handle failing steps and regenerate missing outputs on subsequent runs, without having to run all steps again. For more information about PSOM features refer to Bellec et al. [2012]. The output of the preprocessing pipeline is used for model building and performance evaluation. For more details on pipeline implementation and a description of the individual steps, see Appendix B.

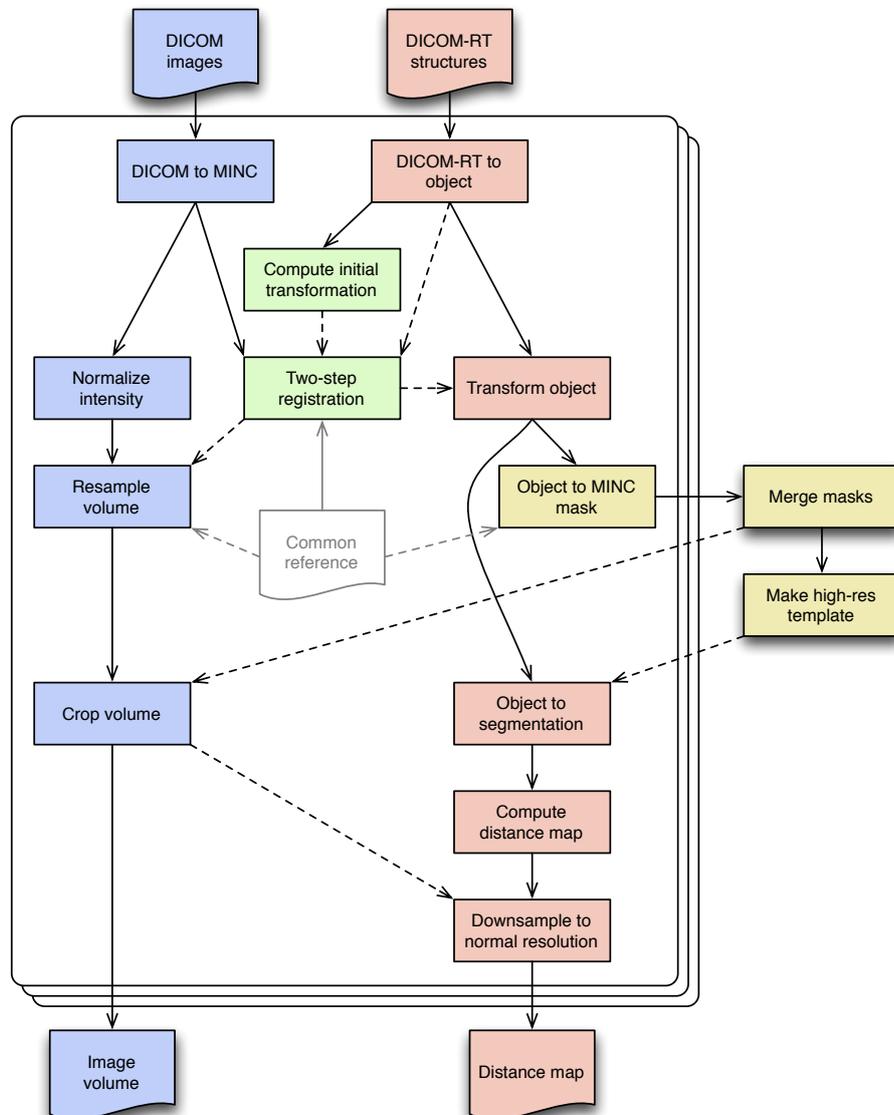


Figure 7.4: Overview of the preprocessing PSOM pipeline for DICOM input (simplified). Steps are color-coded according to: *Image preprocessing*, *Segmentation preprocessing*, *Registration*, and *Cropping*. Solid edges indicate primary input; dashed edges indicate secondary input. Processing steps inside the box are done for each input image, while the three cropping steps outside run only once.

Performance evaluation

The performance of the appearance model-based segmentation algorithm is evaluated using leave-one-out cross-validation, in which the appearance model is trained on all patients but one and then applied for segmentation of the images from the one patient left out. This is done for each patient in turn, and the real performance of the segmentation is then estimated from the median and spread of the individual performance results. By using this validation method the number of training images for model building is maximized, thus maximizing the possible amount of variability described by the model. The validation framework is implemented as a PSOM pipeline (Appendix B).

Training is done using the preprocessed images, but segmentation is done on the original images to assess the segmentation performance in a real clinical setting. The initial parameter estimate is initialized to zero in appearance, rotations, scale, and texture scale. The translations are initialized to the center of the manual segmentation computed from the MNI .obj representation. In a clinical setting the initial translations could be a user-supplied coordinate of the center of the anatomical object in the image.

For each image the following segmentation metrics are computed, comparing the manually delineated segmentation M to the automatic segmentation A :

- Dice kappa
- Jaccard similarity
- Normalized volume difference
- Hausdorff distance
- Median signed distance
- Median absolute distance

The manual segmentations are used as “gold standard” reference, even though the inter- and intra-rater variability can not be assessed.

Dice kappa

The Dice kappa, also known as the Dice similarity coefficient, is a measure of overlap between regions, and is defined as the ratio of the volume in common between the regions to the mean volume of

the regions [Feuerman & Miller, 2008; Zijdenbos et al., 1994; Zou et al., 2004]:

$$\kappa(M, A) = 2 \cdot \frac{V(M \cap A)}{V(M) + V(A)} \quad (8.1)$$

The function $V(R)$ is the volume of a segmented region R .

κ has a value between zero and one, and values close to one implies that the two regions share many common voxels.

Jaccard similarity

The Jaccard similarity coefficient is a measure similar to the Dice kappa, and is defined as the ratio of the common volume to the total volume of segmented voxels:

$$J(M, A) = \frac{V(M \cap A)}{V(M \cup A)} \quad (8.2)$$

Like the Dice kappa a Jaccard similarity coefficient close to one indicates a high level of agreement between the segmentations.

Normalized volume difference

The normalized volume difference is the absolute difference between the volume of the two regions normalized by the mean volume:

$$D(M, A) = 2 \cdot \frac{|V(M) - V(A)|}{V(M) + V(A)} \quad (8.3)$$

A value close to zero indicates that the two regions have roughly the same volume.

Distance metrics

The *Hausdorff distance* is the longest distance between any voxel m in M and any voxel a in A , and is defined as:

$$H(M, A) = \max(h(M, A), h(A, M)) \quad (8.4)$$

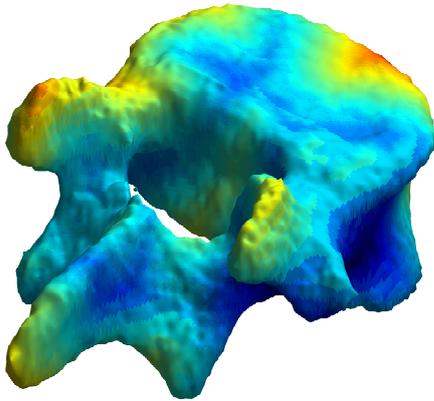
where $h(M, A)$ is the shortest distance from the point m in M , which is farthest away from any point in A , to the point a closest to m :

$$h(M, A) = \max_{m \in M} \min_{a \in A} |m - a| \quad (8.5)$$

Thus all points in one region lie within the Hausdorff distance from a point in the other region [Huttenlocher et al., 1993; Rockafellar & Wets, 1998]. The larger the Hausdorff distance, the more mismatch between the two segmentations.

The Hausdorff distance is estimated using signed distance maps. The signed distance map and MNI .obj representation of each segmentation is computed. The object representation can be computed from

Automatic segmentation on manual distance map



Manual segmentation on automatic distance map

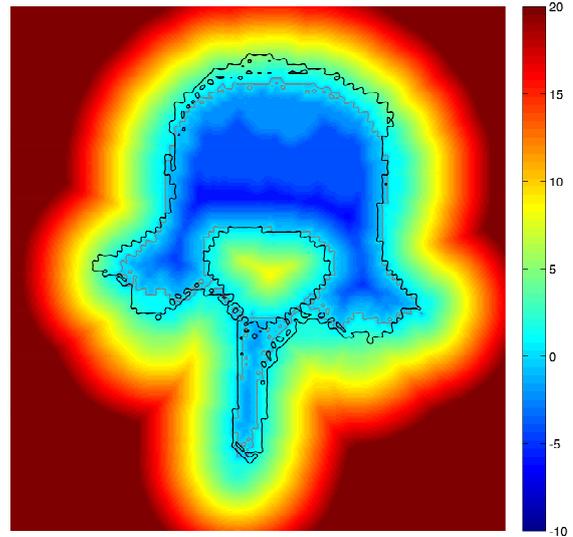


Figure 8.1: Example of the intersection of an object on a distance map. Left: Automatic segmentation colored by the value of the manual distance map. Right: Manual segmentation (black contour) overlaid on the automatic distance map (automatic segmentation shown as gray contour). The value of the distance map on each point of the object surface is the shortest distance from that point to the surface of the other object. Color scale is the same in both figures.

the signed distance map by extracting the zero level-set using marching cubes. These level-sets represent the surface of the segmented objects.

By intersecting the surface of one object with the distance map of the other object the distance to the other object from each point on the surface can be found (Fig. 8.1). The value of the distance map at a point on the surface represents the shortest distance from that point to the other object. By spline interpolation of the distance map to the exact coordinate given by the object representation sub-voxel precision of the distance from the point to the object can be achieved.

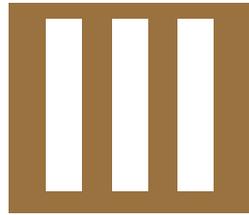
The maximum distance from a point on the surface of one object to the other object is the value $h(\dots)$ in (8.4). An estimate of the Hausdorff distance $H(M, A)$ is found by computing the distances from each object to the other and finding the signed maximum distance. Negative distances represent points on the surface which are inside the other object, and thus the real distance from that point to the other object is zero.

Much of the variability of the vertebrae is in the spinous and transverse processes, which are elongated structures. The Hausdorff distance (or maximum distance) is quite sensitive to variation in such

structures, and does not measure the overall match between the segmentations. The *median signed distance* is less sensitive to variations in such structures, but more sensitive to variation in the overall match. The sign of the distance is defined so that the distance is positive if the automatic segmentation lies inside (is smaller than) the manual reference.

If the objects are roughly the same size, but offset from each other, the median signed distance will be close to zero because the surface of one object will intersect with approximately the same number of positive and negative distances on the signed distance map of the other object. The *median absolute distance* will reveal such a configuration.

PART



Results

Prostate segmentation

An appearance model is trained and tested on 42 MR image volumes of the prostate (chapter 4) in a leave-one-out cross-validation framework (chapter 8). As suggested by Cootes & Taylor [2004] and Hu & Collins [2007] a threshold is applied to remove the least significant eigenvectors from the shape, texture and appearance model. 98% of the variance in each model is retained, reducing the size of the model and the number of parameters [Cootes & Taylor, 2004; Hu & Collins, 2007].

Based on predictions of parameter change learned from the training data set the appearance model is applied for prostate segmentation (chapter 6).

9.1 Appearance model

The appearance model is trained on pre-aligned images. Figure 9.1 gives a visual assessment of the alignment of the training data.

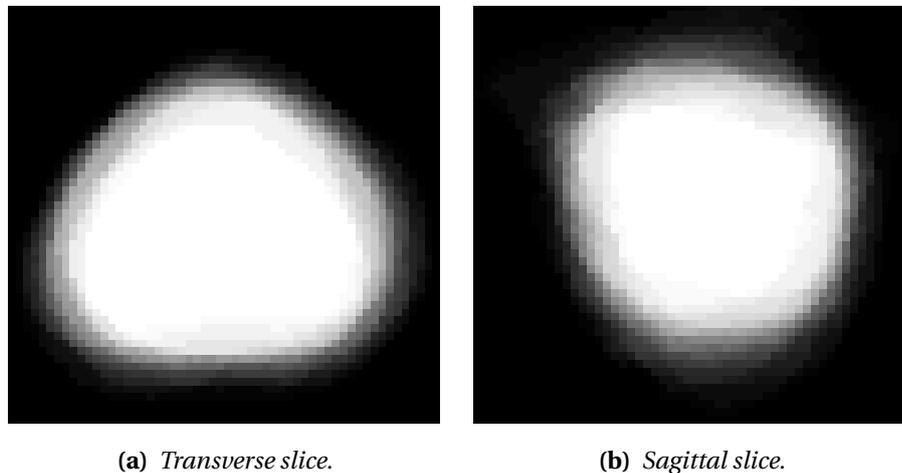


Figure 9.1: *Summation image for visual assessment of the alignment of the prostate training data set.*

To explore the properties of the appearance model it is trained on the full prostate training data set, and the implications of the parameters are examined. The appearance model is a combination of a shape and a texture model (chapter 5). The contribution of each shape and texture parameter to the total variability in each model is shown in Figure 9.2.

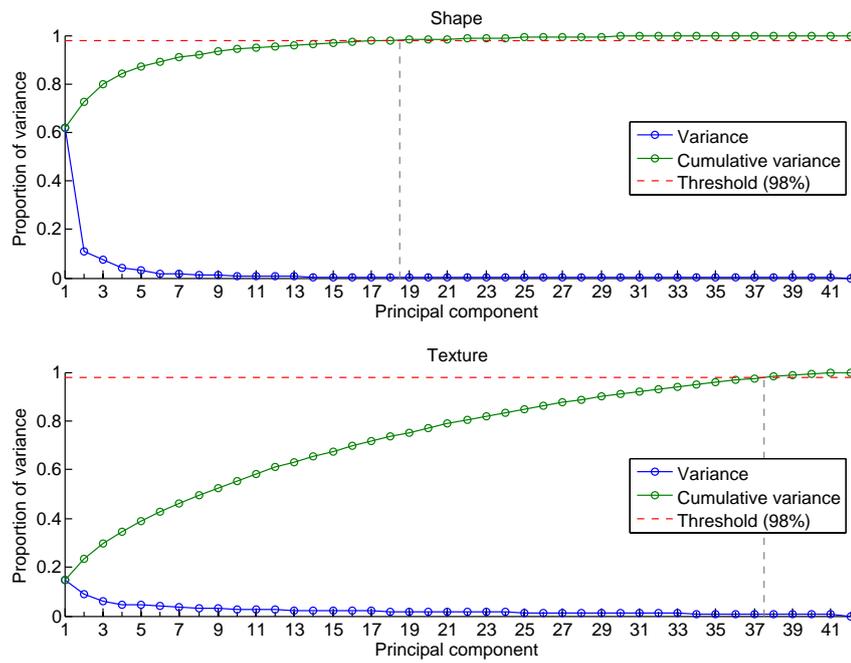


Figure 9.2: Scree plot of shape and texture eigenvalues showing the proportion of variance and cumulative variance for each eigenvector. 18 shape and 37 texture eigenvectors are retained after thresholding.

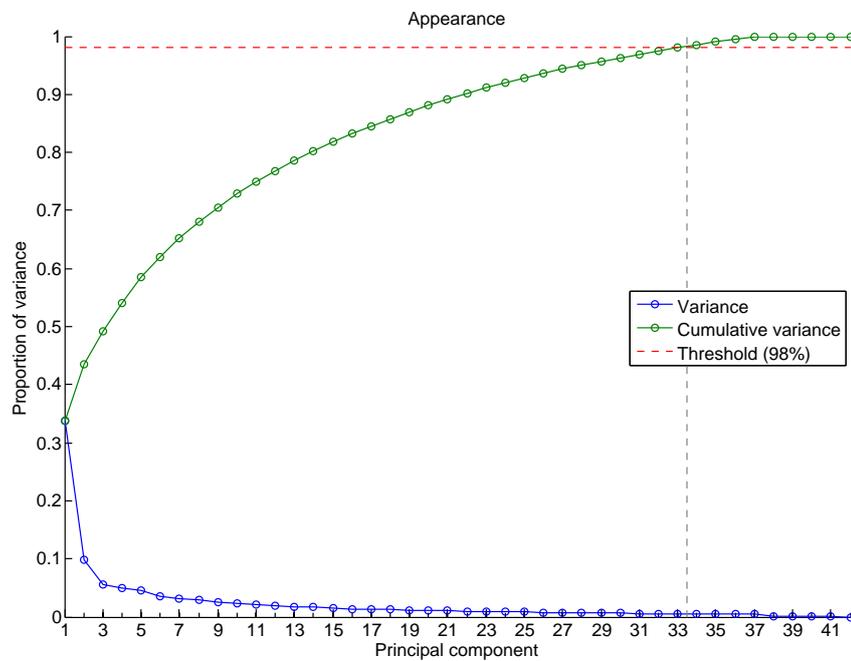


Figure 9.3: Scree plot of appearance eigenvalues. 33 appearance eigenvectors are retained after thresholding.

Only relatively few parameters describe the majority of the variability in the shape model; three parameters describe roughly 80% of the shape variability. After applying the threshold 18 shape parameters are retained.

The variability in the texture model is described by a larger number of parameters. After applying the threshold on the texture model 37 parameters are retained.

The shape and texture models are combined into a full appearance model. The scree plot of the appearance model is shown in Figure 9.3. After applying the threshold 33 appearance parameters are retained. These parameters describe the appearance of the prostate in MR images.

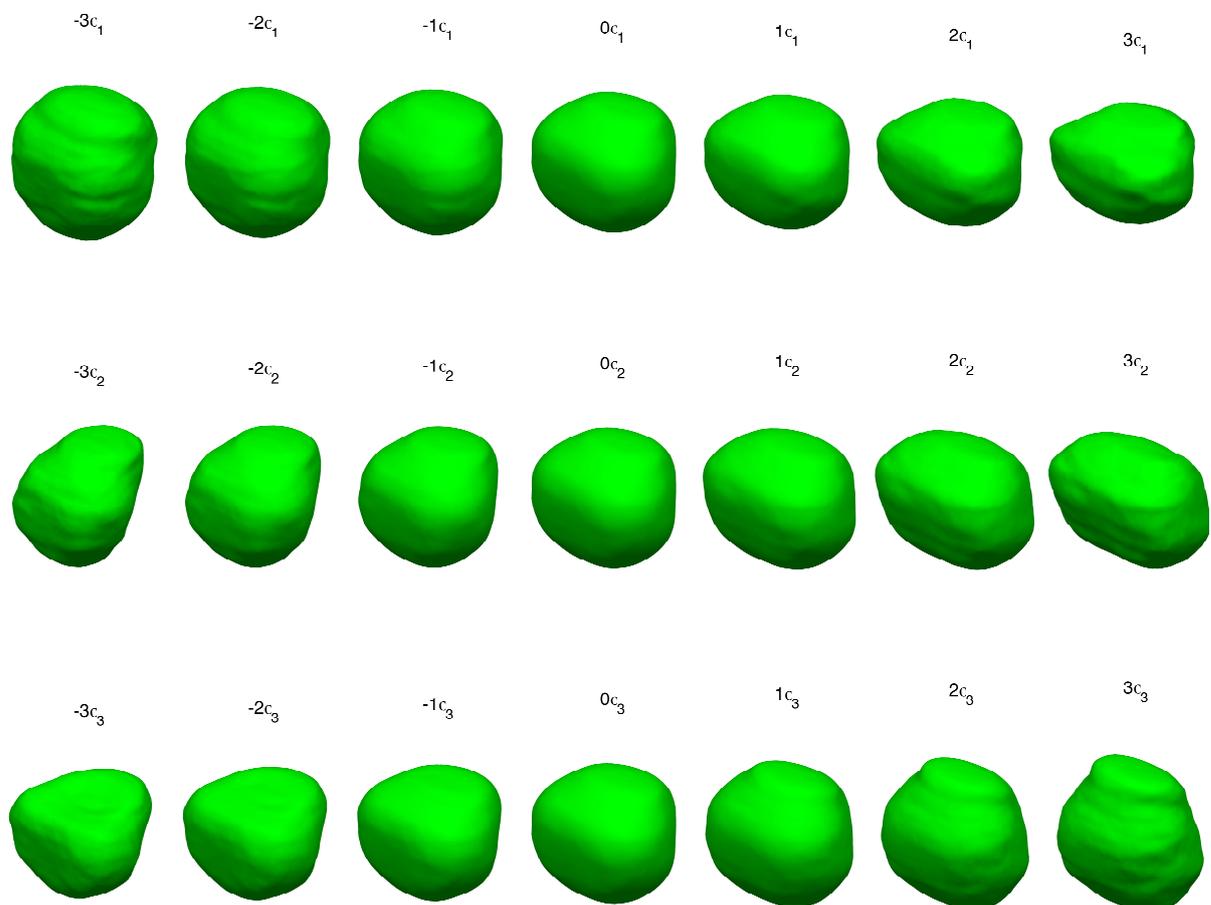


Figure 9.4: Shape variability of the three most significant appearance parameters. Each appearance parameter (rows) is adjusted ± 3 standard deviations (columns).

Figure 9.4 shows the effects on prostate shape of varying the three most significant parameters in the appearance model. Figure 9.5 shows the effects on texture of varying the same parameters. The first parameter corresponds roughly to the overall size of the prostate

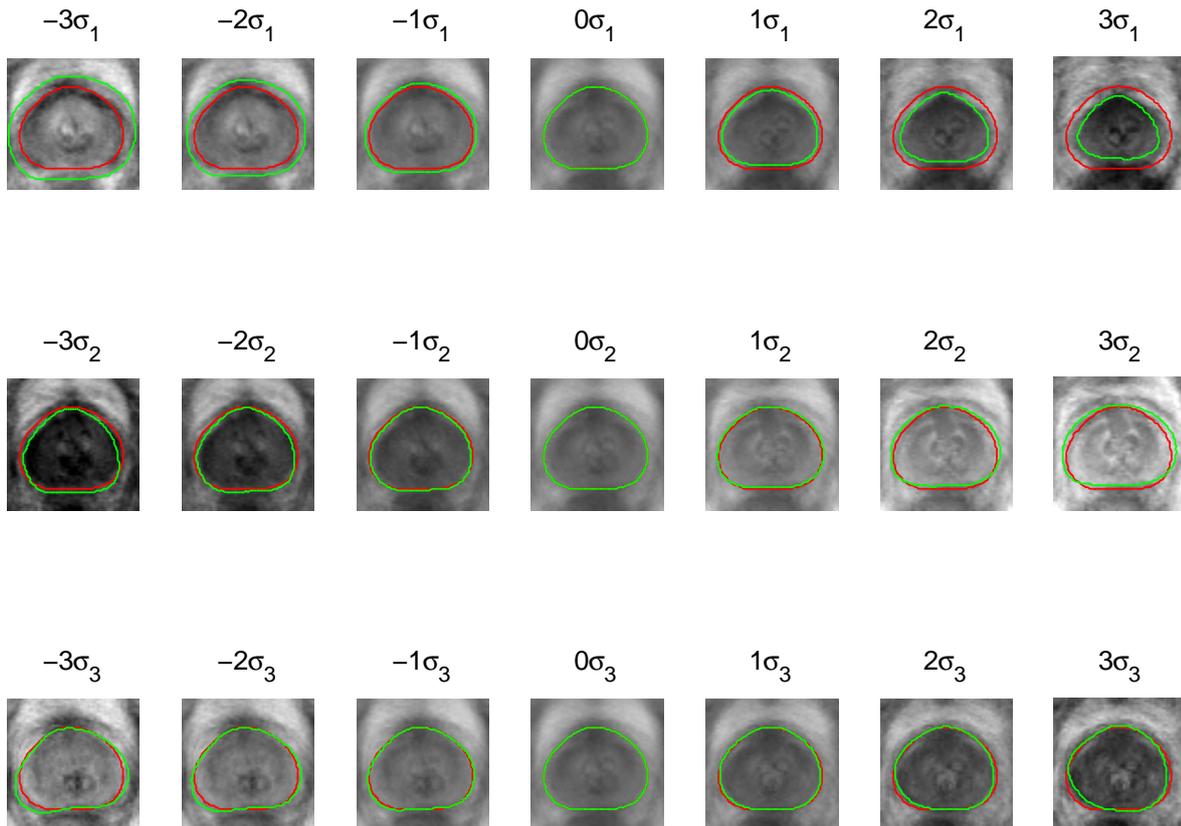


Figure 9.5: *Texture variability of the three most significant appearance parameters in model slice 27. Each appearance parameter (rows) is adjusted ± 3 standard deviations (columns). The model shape is outlined in green, and the mean shape is outlined in red.*

and the ratio between height and diameter in the transverse plane. It also to some extent describes the roundness of the outer surface of the peripheral zone. The second parameter describes the size and extent of the central gland, as well as the length in the lateral directions. The third parameter roughly describes whether the central gland points upwards or forwards, as well as several other more subtle shape changes. The parameters indicate that prostates in the training data set differ most in overall size as well as in the shape of the central gland compared to the peripheral zone.

9.2 Parameter adjustment predictions

To estimate the performance of the parameter adjustment predictions in the iterative segmentation algorithm (chapter 6) each parameter is systematically displaced from the optimal value, and the predicted parameter adjustment is computed based on the difference between the model and image. This is done for each image in the training data set.

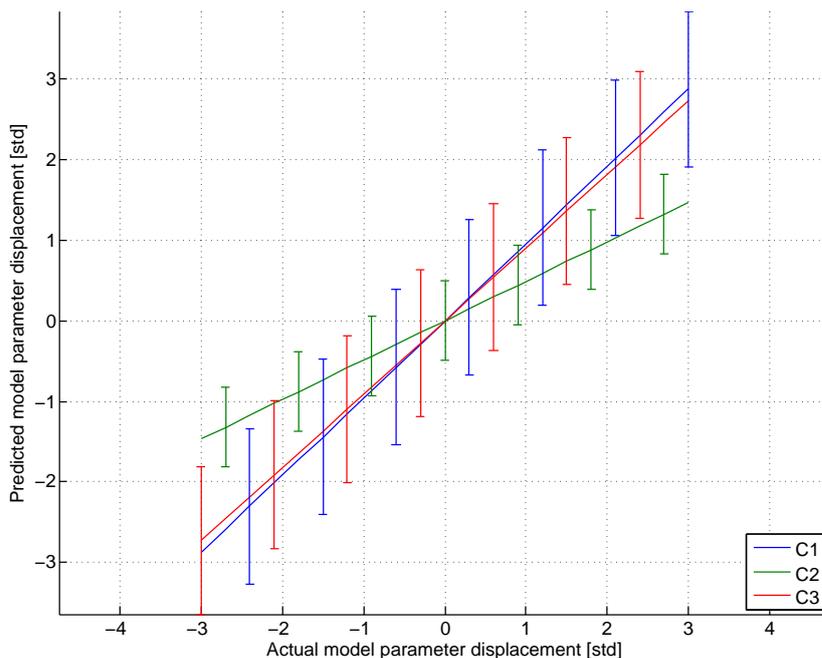


Figure 9.6: Actual vs. predicted displacement of the three most significant prostate appearance parameters. Error bars are one standard deviation.

There is a linear relationship between the predicted adjustment and the actual displacement of the appearance parameters (Fig. 9.6). For rotations there is a linear relationship within approximately $\pm 10^\circ$, and the predictions have the same sign as the displacement within $\pm 60^\circ$ (Fig. 9.7). Translations show a linear relationship between predicted and actual displacements within ± 2 mm, and predictions have the correct sign within displacements of ± 10 mm. For displacements in scale the predictions are weak, but have the correct sign within $\pm 20\%$ change in scale. Texture scale shows a near perfect linear relationship.

9.3 Segmentation performance

Examples of prostate segmentations using the iterative segmentation method (chapter 6) are shown in Figure 9.8 and 9.9. The first figure compares the manual and automatic segmentation on 2-D slices, and the second compares the manually and automatically segmented shape in 3-D. Each segmentation is computed using a model trained on all images except the segmented image.

Figure 9.10 and Table 9.1 show the performance metrics computed in the leave-one-out cross-validation (chapter 8). The outliers in Dice kappa, Jaccard similarity, and median absolute distance are images from the same patient in which the model fail to estimate the correct scale or translations. The outlier in Hausdorff distance is a patient

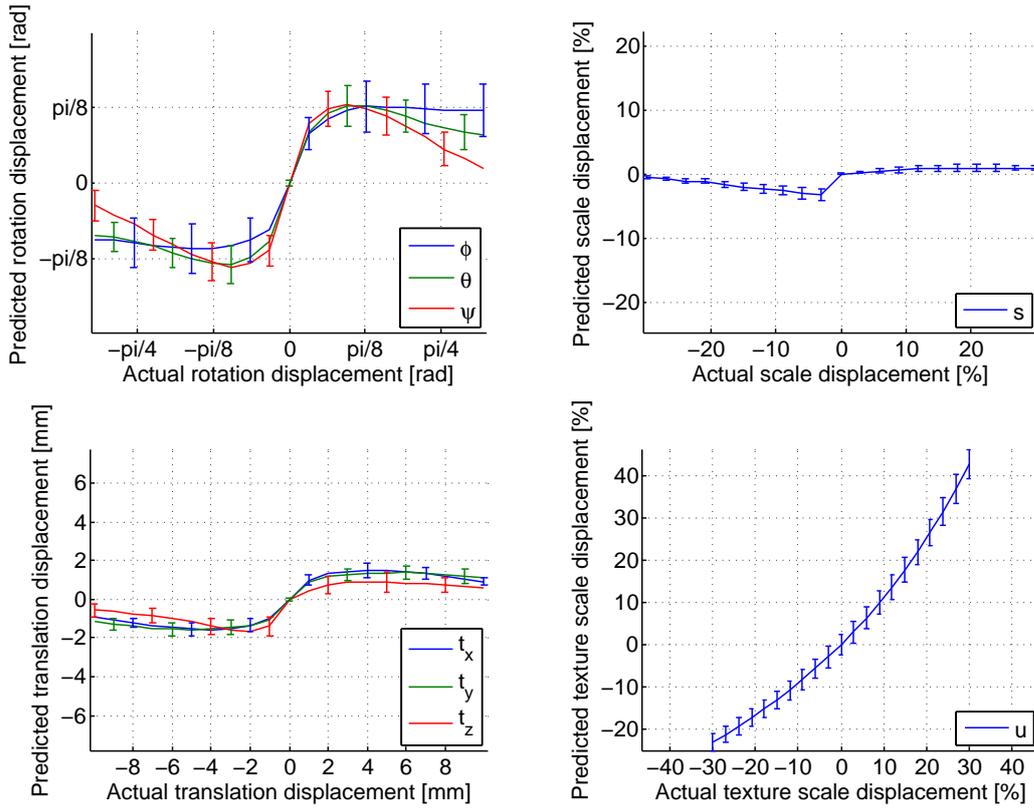


Figure 9.7: Actual vs. predicted displacement of prostate pose and texture parameters. Error bars are one standard deviation.

in which the manual segmentations contain significant amounts of extra-prostatic tissue.

Metric	25 th	50 th	75 th
Dice kappa	0.7435	0.8073	0.8453
Jaccard similarity	0.5917	0.6769	0.7321
Normalized volume difference	0.101	0.2167	0.4613
Hausdorff distance [mm]	6.236	7.528	9.19
Median signed distance [mm]	-1.302	1.63	2.468
Median absolute distance [mm]	1.885	2.375	3.279

Table 9.1: The 25th percentile, the median, and the 75th percentile of the prostate segmentation performance metrics.

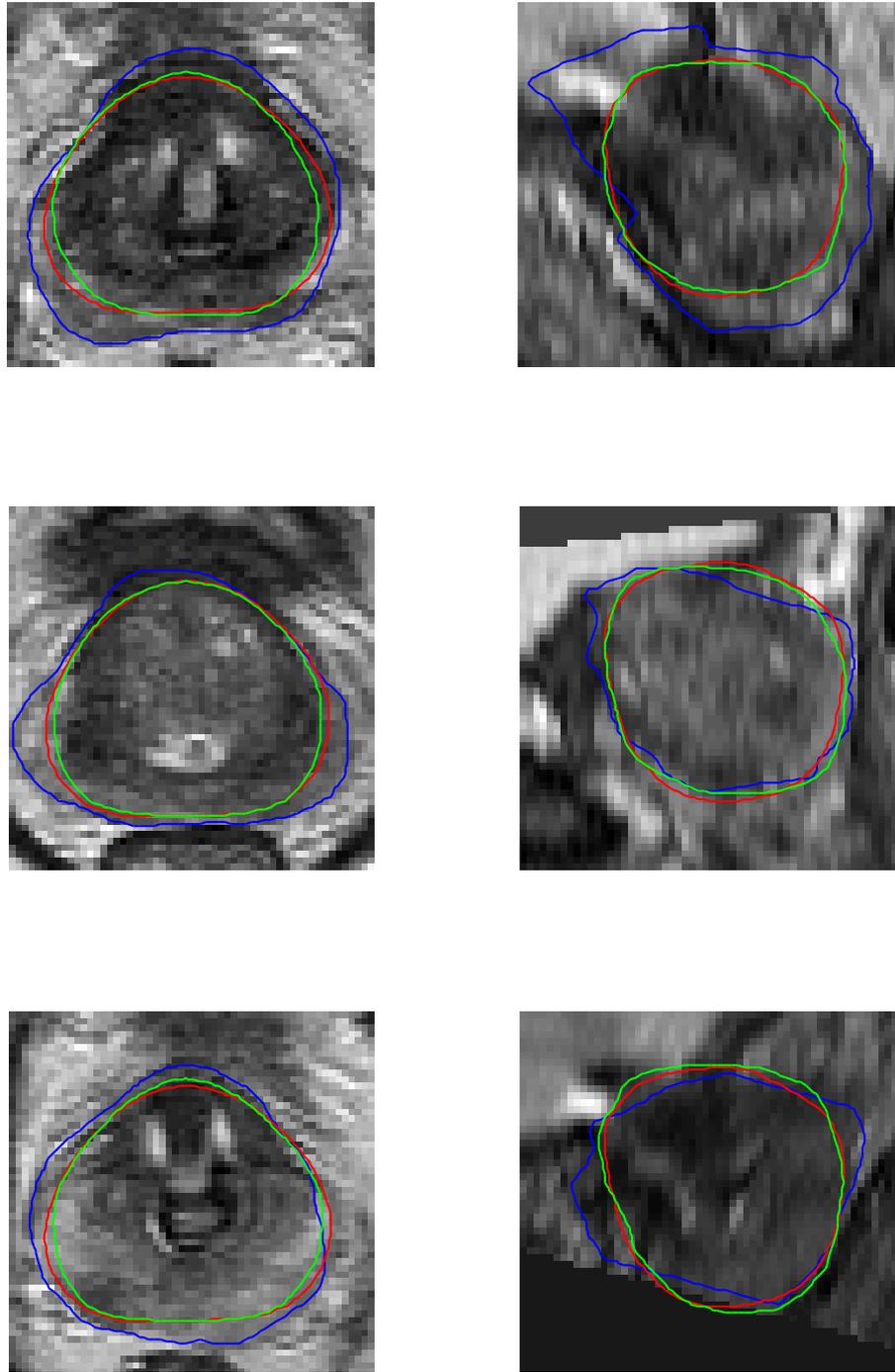


Figure 9.8: Examples of prostate segmentation of scan 1 from patient 2, 3 and 4 respectively (rows). Left column is a transverse slice, right column is a sagittal slice. The automatic segmentation is shown in green, average model in red, and the manual segmentation is shown in blue for reference.

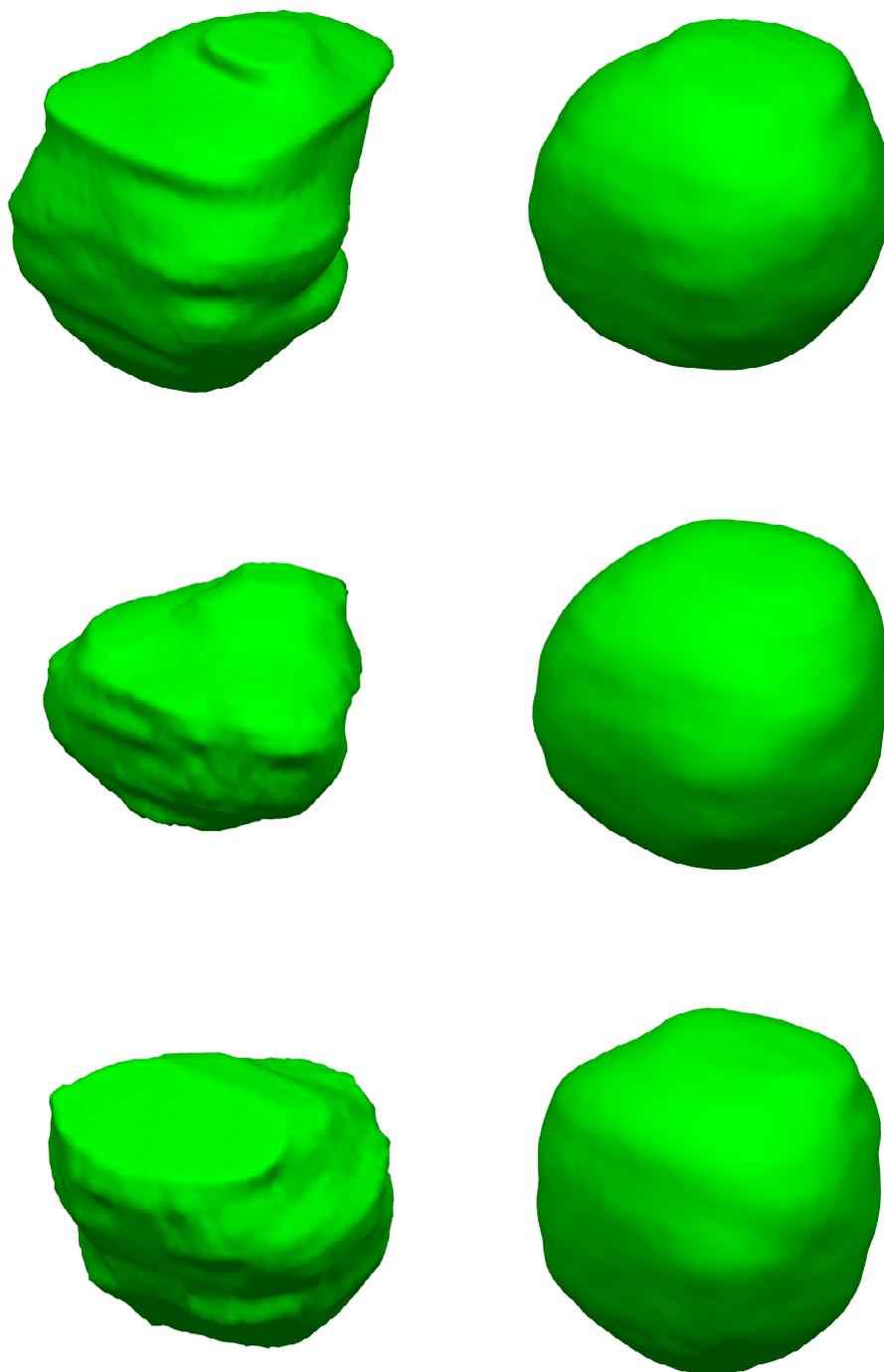


Figure 9.9: *Examples of prostate segmentation of scan 1 from patient 2, 3 and 4 respectively (rows). Left column is manual segmentations, right column is automatic segmentations.*

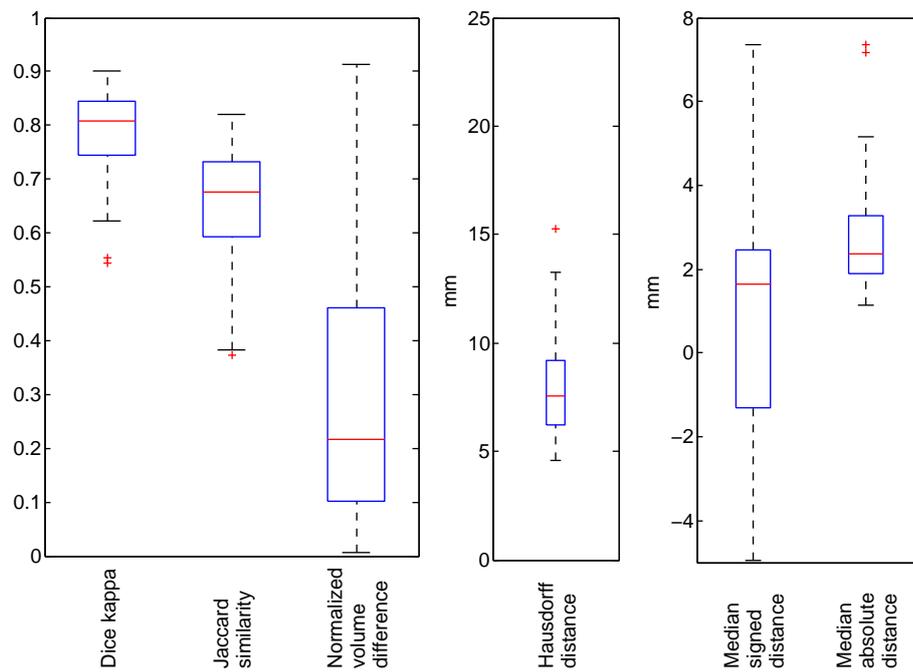


Figure 9.10: Performance metrics computed by leave-one-out cross-validation. Red line is the median, edges of the blue box are the 25th and 75th percentiles, whiskers are the lowest and highest values not considered outliers, and red plusses are outliers. Distance metrics are in units of mm.

Vertebra segmentation

An appearance model is trained and tested on 27 CT image volumes of the L4 vertebra (chapter 4) in a leave-one-out cross-validation framework (chapter 8). The vertebra appearance model is trained using the same parameters as the prostate appearance model (chapter 9).

10.1 Appearance model

The vertebra appearance model is trained on pre-aligned images. Figure 10.1 gives a visual assessment of the alignment of the training data.

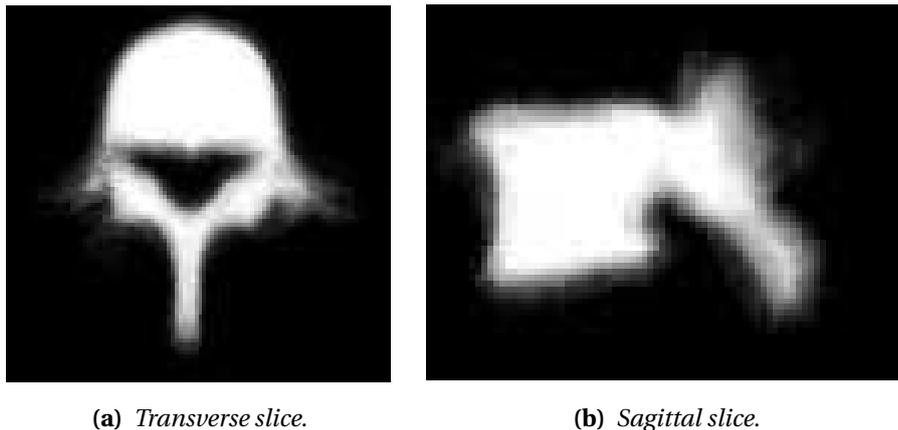


Figure 10.1: *Summation image for visual assessment of the alignment of the vertebra training data set.*

The contribution of each shape and texture parameter to the total variability of the L4 vertebrae is shown in Figure 10.2. A threshold is applied on the parameters to retain 98% of the variance in the model. After applying the threshold 21 shape parameters and 24 texture parameters are retained.

The vertebra shape and texture models are combined into a full appearance model. The scree plot of the appearance model is shown in Figure 10.3. After applying the threshold, 22 appearance parameters are retained in the L4 vertebra model.

Figure 10.4 shows the effect on vertebra shape of varying the four most significant appearance parameters. Figure 10.5 shows the effects on texture of varying the same parameters. The first parameter

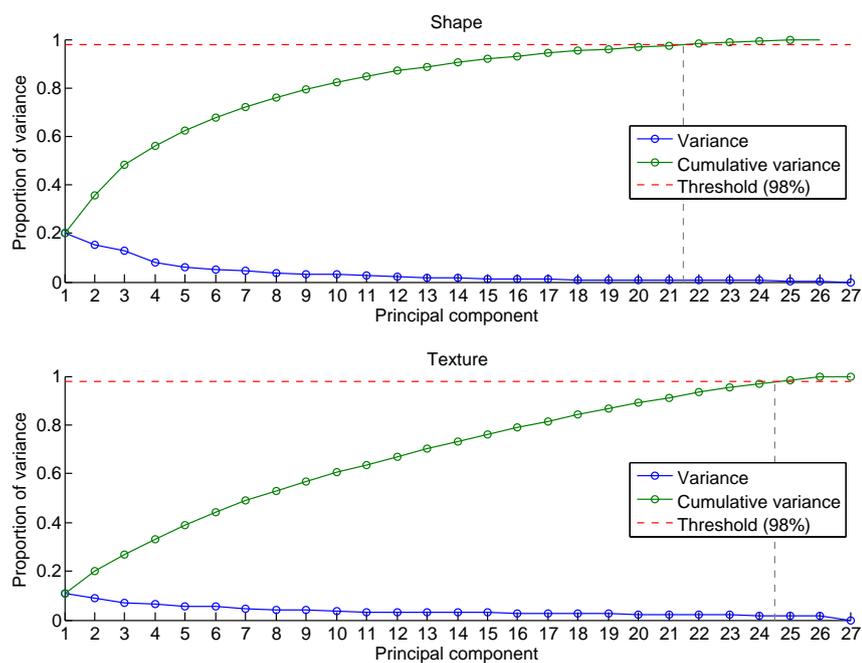


Figure 10.2: Scree plot of shape and texture eigenvalues showing the proportion of variance and cumulative variance for each eigenvector. 21 shape and 24 texture eigenvectors are retained after thresholding.

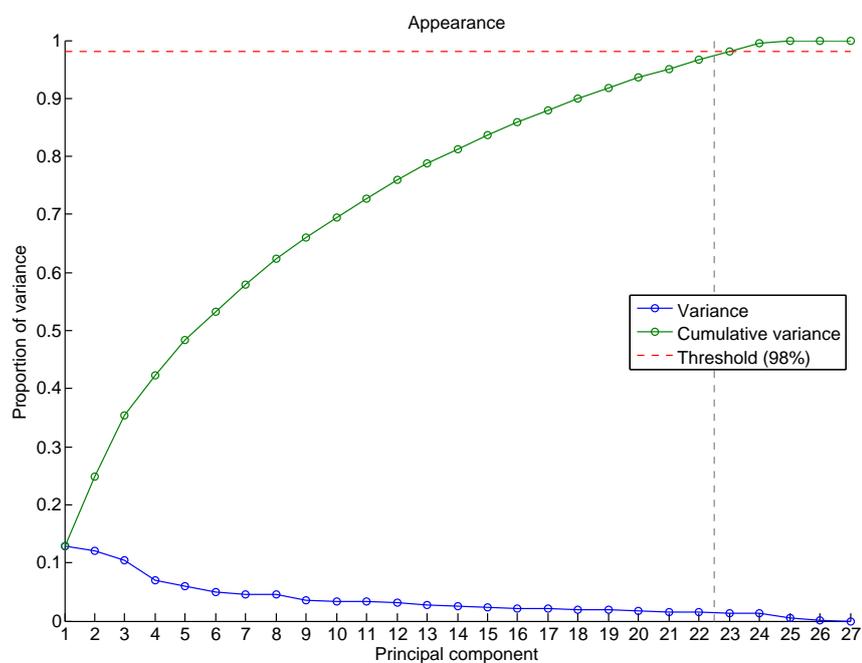


Figure 10.3: Scree plot of appearance eigenvalues. 22 appearance eigenvectors are retained after thresholding.



Figure 10.4: Shape variability of the four most significant appearance parameters. Each appearance parameter (rows) is adjusted ± 2 standard deviations (columns).

is roughly related to the length and shape of the spinous process as well as the upper facet joints. The second parameter is related to the length of the transverse processes and the shape of the lower facet joints. The third parameter is related to the length of the spinous and transverse processes compared to the size of the vertebral body. The fourth parameter is related to the height of the vertebral body in relation to the length of the spinous process.

The parameters indicate that the L4 vertebrae in the training data set differ most in length of the spinous process, shape of the facet joints and length of the transverse processes.

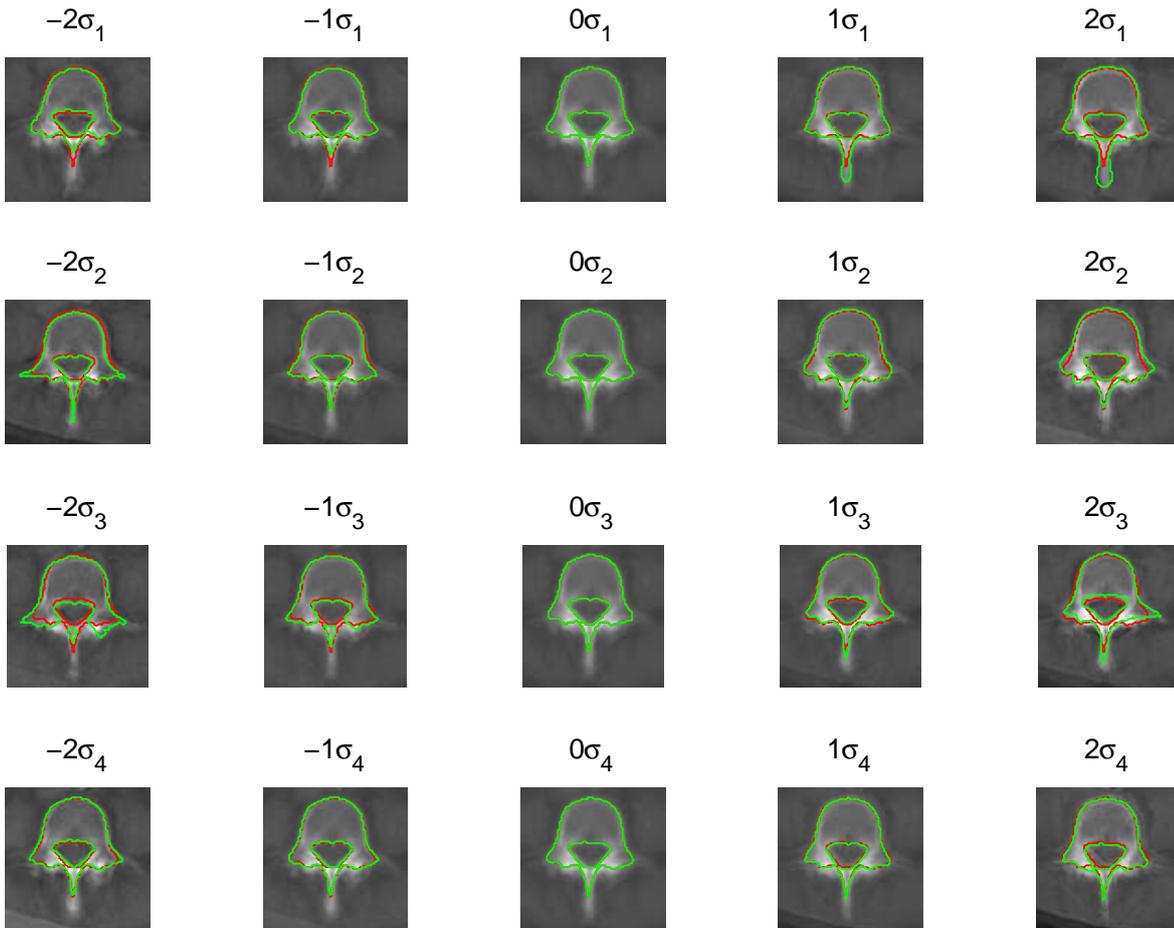


Figure 10.5: *Texture variability of the four most significant appearance parameters in model slice 35. Each appearance parameter (rows) is adjusted ± 2 standard deviations (columns). The model shape is outlined in green, and the mean shape is outlined in red.*

10.2 Parameter adjustment predictions

The relationship between parameter adjustments predicted by the algorithm and the true displacements is shown in Figure 10.6 for appearance parameters and in Figure 10.7 for pose and texture parameters. The relationship between the actual and predicted parameter displacements for the vertebra model is similar to the prostate model (section 9.2).

10.3 Segmentation performance

Examples of vertebra segmentations using the iterative segmentation method (chapter 6) are shown in Figure 10.8 and 10.9, comparing the automatic segmentation to the manual. Each segmentation is computed using a model trained on all images except the segmented image.

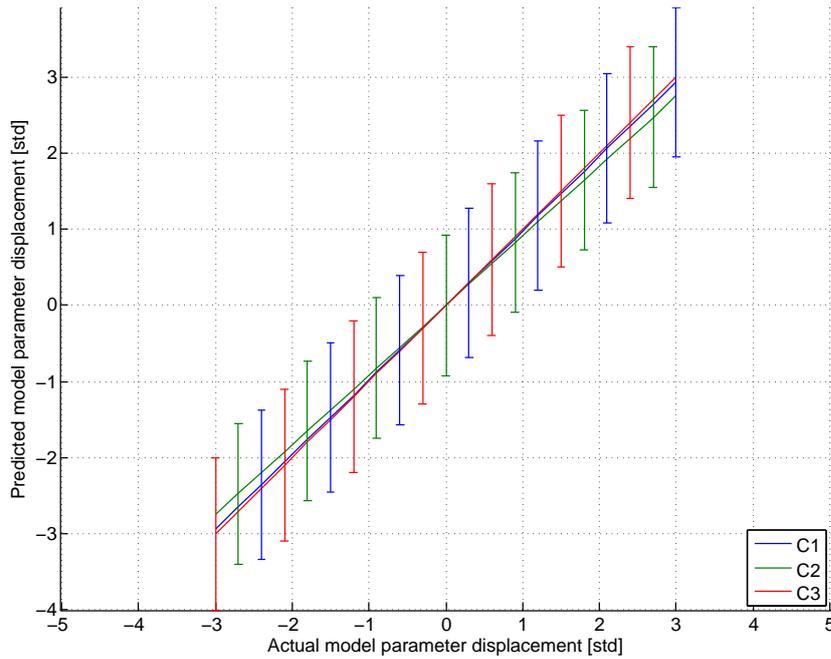


Figure 10.6: Actual vs. predicted displacement of the three most significant vertebra appearance parameters. Error bars are one standard deviation.

Figure 10.10 and Table 10.1 show the performance metrics computed in the leave-one-out cross-validation (chapter 8). The outliers in Dice kappa, Jaccard similarity, and distance metrics are caused by an image in which the model fail to estimate the correct translations, because the initial translation estimates are too far from the real position of the vertebra.

Metric	25 th	50 th	75 th
Dice kappa	0.7816	0.8086	0.8265
Jaccard similarity	0.6415	0.6788	0.7043
Normalized volume difference	0.2285	0.2949	0.3356
Hausdorff distance [mm]	11.37	13.06	15.67
Median signed distance [mm]	1.832	1.93	2.414
Median absolute distance [mm]	1.875	1.978	2.414

Table 10.1: The 25th percentile, the median, and the 75th percentile of the vertebral segmentation performance metrics.

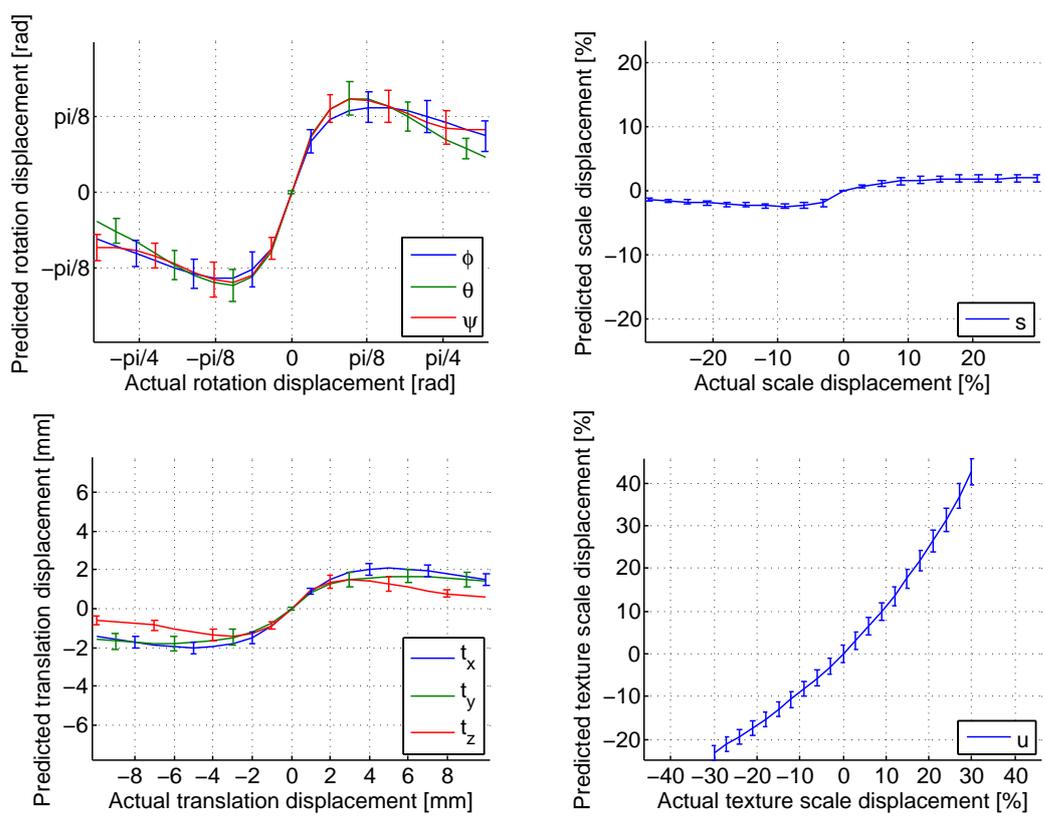


Figure 10.7: Actual vs. predicted displacement of vertebra pose and texture parameters. Error bars are one standard deviation.

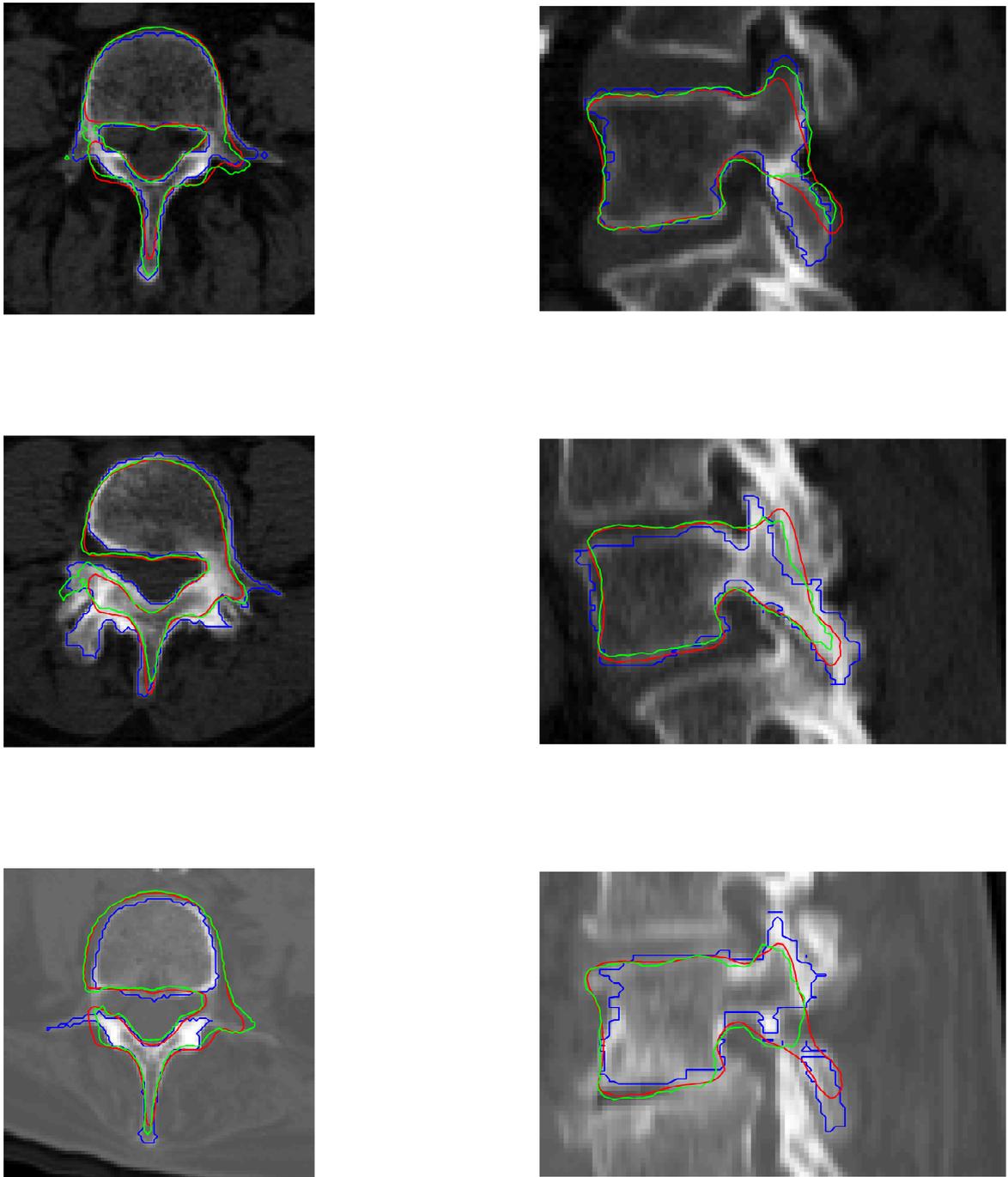


Figure 10.8: Examples of vertebra segmentation from patient 2, 3 and 4 respectively (rows). Left column is a transverse slice, right column is a sagittal slice. The automatic segmentation is shown in *green*, average model in *red*, and the manual segmentation is shown in *blue* for reference.

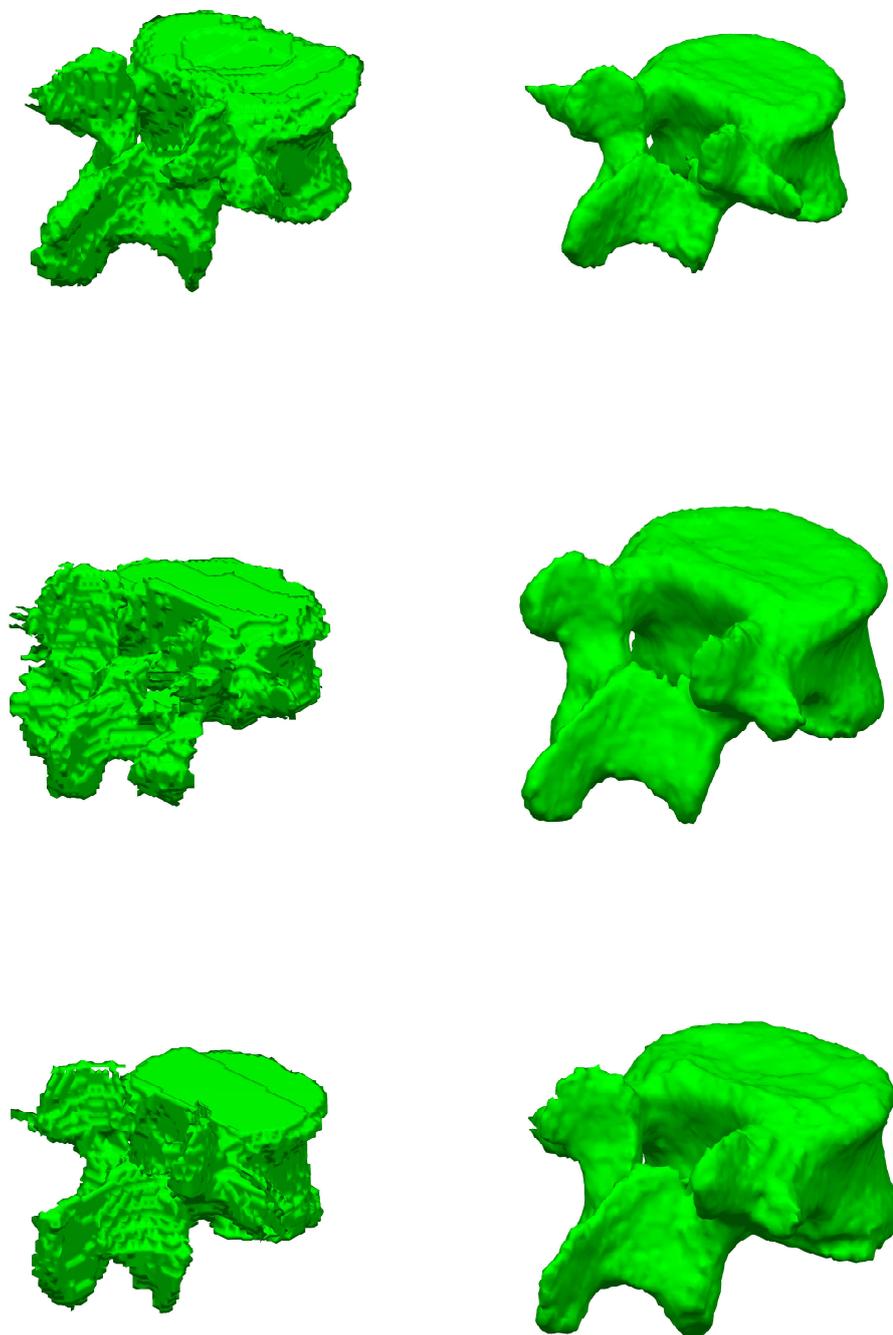


Figure 10.9: *Examples of vertebra segmentation from patient 2, 3 and 4 respectively (rows). Left column is manual segmentations, right column is automatic segmentations.*

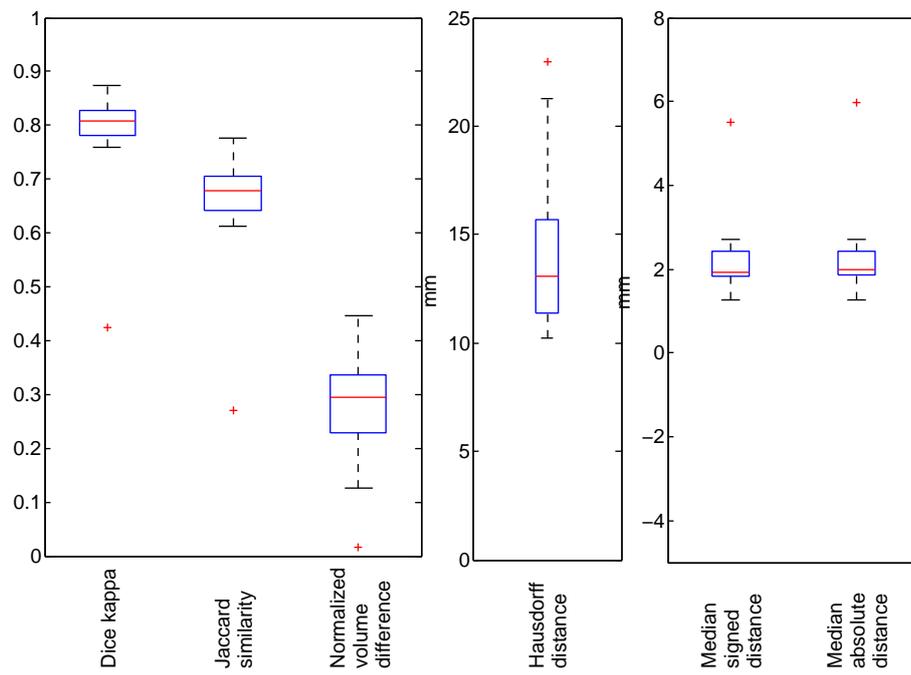


Figure 10.10: Performance metrics computed by leave-one-out cross-validation. Red line is the median, edges of the blue box are the 25th and 75th percentiles, whiskers are the lowest and highest values not considered outliers, and red plusses are outliers. Distance metrics are in units of mm.

PART



Discussion

Discussion

An active appearance model with level-set representation of shape has been developed for 3-D segmentation of medical images in an iterative segmentation algorithm. The appearance model was based on the work by Cootes et al. [1998], which has been extended with a level-set description of shape by Hu & Collins [2007]. Cootes et al. [1998] implemented an iterative segmentation framework in 2-D based on prior knowledge of parameter adjustments, which also incorporated pose changes. Hu & Collins [2007] implemented a recursive least squares search for segmentation in 3-D, but without incorporating pose changes.

In this project the active appearance model with level-set shape representation by Hu & Collins [2007] was utilized in the iterative segmentation framework originally suggested by Cootes et al. [1998] to allow for pose changes during segmentation. The segmentation framework was extended to 3-D and tested on two different anatomical structures.

Data

The modeling and segmentation algorithm was tested on prostate MRI data and vertebra CT data. Original manual segmentations were represented as slice contours and voxel-based binary masks respectively.

Depending on the representation of the manual segmentations different kinds of segmentation errors are seen. It can be clearly distinguished from a 3-D rendering of a segmentation which representation has been used. A segmentation based on slice contours will appear layered with continuous curves in the slice plane, but with a more irregular outline orthogonal to the slice plane. A voxel-based segmentation will easily appear jagged with spurious spikes and holes, especially if the segmentation is not carefully done, which is a difficult task in three dimensions. The goal of modeling is to describe the variability in the underlying anatomical structure and to exclude the segmentation errors.

In the lack of a true gold standard the manual segmentations were used for training and evaluation of the segmentation performance. The segmentation errors mentioned above affected the outcome of the performance metrics. If the appearance model is to make perfect correspondence with the manual segmentations the model must

include the segmentation errors, which is not desirable. The model should exclude the segmentation errors and only describe the underlying structure. Thus a good appearance model will achieve lower than ideal performance metrics when compared to the manual segmentations. The model should achieve better performance metrics if compared to true gold standard segmentations.

This could be dealt with in several ways. A gold standard could be established by evaluating the inter- and intra-rater variability of the manual segmentations, and by computing reference segmentations based on a set of segmentations of the same image. This would, however, require a lot of manual labor to segment each image several times. Another way to deal with the problem is to use simulated data, in which the gold standard is known. Such simulated data could be generated from the model with additive noise as suggested by Betrouni et al. [2011].

Appearance modeling

The parameters of the appearance models are related to the physical properties and anatomical variability of the modeled object. Prostates have a relatively simple shape, and can thus be described by relatively few parameters. The main variability is in size, possibly caused by age-related prostate hypertrophy.

Vertebrae have a more complex shape with several articulations (e.g. spinous and transverse processes) varying in length and size relative to each other. This is also seen in the vertebra appearance model, where a larger ratio of parameters are retained after thresholding. The overall size of the L4 vertebra is more constant in adults than the prostate size.

The threshold applied on the model parameters determine the amount of variability in the training data set to include in the model. If all variability is included the training data will be perfectly described, but the model will over-fit to the errors in the manual segmentations described above. This will result in lower segmentation performance compared to gold standard segmentations. If too few parameters are retained the model will not be able to sufficiently describe the natural variability of the object, again resulting in lower segmentation performance.

In this project the threshold was chosen at 98%, but the threshold could also be determined by evaluating the segmentation performance for models trained at different thresholds and selecting the threshold resulting in the best segmentation performance. However, the training phase is a slow and computationally expensive task, and training the model at different thresholds will take quite some time (approximately one week with the current MATLAB implementation, depending on the number of thresholds to test).

Parameter adjustment predictions

The segmentation algorithm utilizes prior knowledge of how to adjust the model parameters in order to achieve a better fit of the model to the target image. The learned information is an approximation of the relationship between changes in the grayscale intensity residuals and changes in the model parameters.

Ideally there is a linear relationship between the parameter adjustments predicted by the algorithm and the actual displacement from the optimal parameter values. For model and texture parameters the relationship is close to linear. For rotations and translations the relationship is linear for small parameter displacements, but the linear relationship breaks down for larger displacements. However, as long as the predicted and actual parameter displacements have the same sign and the algorithm does not over-predict too far it should still converge [Cootes et al., 1998]. For scaling the linear relationship is almost non-existing, which results in segmentation failure if the scale of the object in the target image is too far from the mean object scale. The relationships between predicted and actual parameter displacements are similar for the prostate and vertebra model, and are also consistent with the relationships found by Cootes et al. [1998] for a 2-D appearance model of human faces.

To improve the robustness to large displacements in pose parameters Cootes et al. [1998] suggest to implement a multi-resolution algorithm, in which the model is trained at different resolutions in a Gaussian image pyramid. During segmentation the model would then be fitted to the target image starting at the lowest resolution to achieve a coarse fit for large pose displacements, continuing to finer resolutions to refine the model fit. This could be the focus of future work.

Segmentation algorithm

The segmentation algorithm is a rather simple iterative algorithm similar to gradient descent, with the difference that the gradient is estimated using prior knowledge and not recalculated at every step. To allow for different step sizes in the gradient descent a simplistic line search is applied, in which different step sizes are tried in turn until a better fit is found. This might not be the most efficient approach, and if speed is an issue it would be relevant to investigate the effect of using other optimization methods (e.g. golden section search [Kiefer, 1953] or Brent's method [Brent, 1973]) in the line search.

The current MATLAB implementation is able to segment an image in about 15 seconds, which may be fast enough for typical clinical applications.

Prostate segmentation results

Prostate segmentation results show a median Dice kappa of 0.81, which is lower than most previously reported results. A previous

atlas-based method achieved a mean Dice kappa of 0.85, but the results were achieved using data with thinner slices than in the current study (1 mm), and the algorithm was significantly slower than the algorithm presented in this study (15 min compared to 15 sec in this study) [Klein et al., 2008]. A landmark-based appearance model achieved a Dice kappa of 0.88 ± 0.06 using Haar wavelets for extracting texture features [Ghose et al., 2010]. This approach could also be applied in the current appearance model instead of using the grayscale intensity values as texture features directly. The segmentation results are similar to a level-set shape model achieving a Dice kappa of 0.83 ± 0.06 by fitting the model to binary images extracted by voxel classification [Chowdhury et al., 2012]. The median Jaccard similarity of 0.68 is also lower than a previously reported mean Jaccard coefficient of 0.78 [Pasquier et al., 2007] achieved using active shape models, but that method was more an aid in manual segmentation in that it applied manual re-initialization or corrections of erroneous contours. Also the method was quite slow with an estimated segmentation time of 1 hour per patient, which was only in some cases faster than manual segmentation.

The median signed distance of 1.6 mm show that the automatic prostate segmentation is in general smaller than the manual reference, indicating that the algorithm fails to determine the correct prostate size. The larger median absolute distance of 2.4 mm indicate that the two segmentations are also somewhat misaligned. This result is similar to a previous shape model-based segmentation achieving a mean absolute distance of 2.6 ± 0.6 mm [Chowdhury et al., 2012] and an appearance model-based segmentation achieving a mean absolute distance of 3.97 ± 2.74 mm [Ghose et al., 2010]. Outliers in the results are cases in which the algorithm fail to determine the correct scale and/or translations. These findings are consistent with the comparison of actual vs. predicted pose parameter displacements in that the scale prediction is very weak and the translation predictions break down for large displacements.

The results show that some scale is described by the appearance model, and some scale is described by the scale parameter. To remove scale completely from the appearance model the initial registration of the training data should be improved. Another approach is to include scale completely in the appearance model and discard the scale parameter from pose. This implies the assumption that the prostate size varies in a way that can be described statistically, which may be a reasonable assumption.

It should be noted that the distance maps are computed from a voxelated segmentation representation, introducing an uncertainty of ± 0.1 mm in the distance map for manual prostate segmentations and ± 0.5 mm for automatic segmentations.

The median Hausdorff distance of 7.5 mm indicate that the model fails to describe protrusions in the prostate shape. Looking at the manual reference it is evident that these protrusions are extra-prostatic tissue included in the clinical target volume, which should probably not be included in gold standard prostate segmentations. This is especially the case for the single outlier in Hausdorff distance. The model fails to represent these structures, because they are not sufficiently represented in the training data set. This is also shown by the 3-D example renderings of manual and automatic segmentations, where the automatic segmentations are closer to the mean model shape than the manual segmentations. For use in radiotherapy planning it is important that the Hausdorff distance is as small as possible to ensure that the complete prostate volume receive full radiation dose.

Vertebra segmentation results

Vertebra segmentations show results similar to the prostate segmentations. The median Dice kappa is 0.81, the Jaccard coefficient is 0.68, and the normalized volume difference is 0.29, but all three metrics have a lower spread than for prostate segmentations. No authors have reported these metrics for full vertebra segmentations in the literature, thus a direct comparison with other methods is not possible.

The outlier in Dice kappa and Jaccard similarity coefficient is a case in which the initial translation guesses are too far from the actual vertebra location causing the algorithm to fail to converge to the correct location.

The lower spread in metrics are probably due to the fact that the manual references contain only the vertebra with no surrounding tissue, as in the prostate data set. Another factor could be that the vertebra shape is more distinct than the prostate, and that the grayscale intensity contrast with the background is higher.

The median signed distance and median absolute distance are both very similar with a median value of 1.9 mm, indicating that the segmentations are well aligned (except for the mis-aligned outlier), but that the automatic segmentations are 1–2 mm smaller than the manual references. The median Hausdorff distance of 13 mm indicate that the model fails to segment the vertebral protrusions correctly. Note that the distance map uncertainty introduced by the voxelated segmentation representation is ± 0.5 mm for both manual and automatic vertebra segmentations.

The 3-D example renderings of automatic and manual segmentations show that the model describes less variability than what is present in the manual segmentations. This is actually an advantage of the vertebra appearance model, because the manual segmentations contain quite a bit of shape “noise”, i.e. errors introduced by the

voxel-based manual segmentation method, but it also results in less accurate segmentations of the vertebral protrusions. An appearance model trained on more data will contain more shape variability, while at the same time be able to exclude the manual segmentation errors from the model.

11.1 Conclusion and directions for future work

The project has shown that it is possible to build an active appearance model based on clinical manual segmentations with fair results, but also that validation requires true gold standard reference segmentations, in which segmentation errors and the inter- and intra-rater variability are minimized. The active appearance model can be applied on any structure of interest if the structure shape is not too variable and varies in a statistical manner. The segmentation results are slightly poorer than in previous studies applying similar shape or appearance model based-methods, but a true comparison is difficult because different data sets are used, and because different degrees of manual intervention are involved.

The results show that the segmentation algorithm must be more robust to variation in size and to translation displacements caused by erroneous initial pose estimates. This could possibly be achieved by implementing a multi-resolution segmentation algorithm as suggested by Cootes et al. [1998].

In radiotherapy dose planning for treatment of prostate cancer it is relevant to expand the method for segmentation of surrounding structures like the seminal vesicles as well. This is straightforward to do with the current shape representation, and it has been shown using a shape model by Tsai et al. [2004].

An interesting question is how the size of the training data set affects the segmentation performance. Intuitively more variable structures require more training data to be fully described by a statistical model, but it is not straightforward to determine the sufficient amount of training data for a given structure. The current and previous studies seem to indicate that statistical models perform reasonably well with small training sets because the mean shape is quite close to the actual shape of the object of interest, but also that large training sets are required for the models to perform as well as more specialized segmentation methods.

Bibliography

- Arand, M., Hartwig, E., Kinzl, L., & Gebhard, F. (2002, June). Spinal navigation in tumor surgery of the thoracic spine: First clinical results. *Clinical Orthopaedics and Related Research*, 399, 211–218.
- Aslan, M., Ali, A., Chen, D., Arnold, B., Farag, A., & Xiang, P. (2010, September). 3D vertebrae segmentation using graph cuts with shape prior constraints. In *Image processing (ICIP), 2010 17th IEEE international conference on* (pp. 2193–2196). IEEE. doi: 10.1109/ICIP.2010.5652849
- Aslan, M., Ali, A., Farag, A., Abdelmumin, H., Arnold, B., & Xiang, P. (2011). A new segmentation and registration approach for vertebral body analysis. In *Biomedical imaging: From nano to macro, 2011 IEEE international symposium on* (pp. 2006–2009). IEEE. doi: 10.1109/ISBI.2011.5872805
- Aslan, M., Ali, A., Farag, A., Rara, H., Arnold, B., & Xiang, P. (2010, August). 3D vertebral body segmentation using shape based graph cuts. In *Pattern recognition (ICPR), 2010 20th international conference on* (pp. 3951–3954). IEEE. doi: 10.1109/ICPR.2010.961
- Becker, M., Kirschner, M., Fuhrmann, S., & Wesarg, S. (2011). Automatic construction of statistical shape models for vertebrae. In G. Fichtinger, A. Martel, & T. Peters (Eds.), *Medical image computing and computer-assisted intervention – MICCAI 2011* (Vol. 6892, pp. 500–507). Springer Berlin / Heidelberg. doi: 10.1007/978-3-642-23629-7_61
- Bellec, P. (2011a). *Neuroimaging analysis kit for matlab and octave (NIAK)*. Retrieved from <http://code.google.com/p/niak/>
- Bellec, P. (2011b). *Pipeline system for octave and matlab (PSOM)*. Retrieved from <http://code.google.com/p/psom/>
- Bellec, P., Lavoie-Courchesne, S., Dickinson, P., Lerch, J., Zijdenbos, A., & Evans, A. C. (2012, April). The pipeline system for octave and matlab (PSOM): a lightweight scripting framework and execution engine for scientific workflows. *Frontiers in Neuroinformatics*, 6(7), 1–18. doi: 10.3389/fninf.2012.00007
- Benameur, S., Mignotte, M., Parent, S., Labelle, H., Skalli, W., & de Guise, J. (2003). 3D/2D registration and segmentation of scoliotic vertebrae using statistical models. *Computerized Medical Imaging and Graphics*, 27(5), 321–337. doi: 10.1016/S0895-6111(03)00019-3

- Betrouni, N., Iancu, A., Puech, P., Mordon, S., & Makni, N. (2011). Prostatlas: A digital morphologic atlas of the prostate. *European Journal of Radiology*. doi: 10.1016/j.ejrad.2011.05.001
- Borgefors, G. (1991). Another comment on “a note on ‘distance transformations in digital images’”. *CVGIP: Image Understanding*, 54(2), 301–306. doi: 10.1016/1049-9660(91)90070-6
- Borre, M., Iversen, P., Bendixen, A., Iversen, M. G., & Kehlet, H. (2008, August). Organisation og tidlige operationsresultater efter radikal prostatektomi i Danmark 2004–2007. *Ugeskrift for Læger*, 170(34), 2545–2549.
- Brasso, K. (2007, May). Prostatcancer – forekomst og risikofaktorer. *Ugeskrift for Læger*, 169(20), 1883–1886.
- Brent, R. (1973). Algorithms for minimization without derivatives. In (chap. 4). Prentice-Hall.
- Counce, A., & Taylor, C. (1999). Using local geometry to build 3D sulcal models. In A. Kuba, M. Šámal, & A. Todd-Pokropek (Eds.), *Information processing in medical imaging* (Vol. 1613, pp. 196–209). Springer Berlin / Heidelberg. doi: 10.1007/3-540-48714-X_15
- Chowdhury, N., Toth, R., Chappelow, J., Kim, S., Motwani, S., Punekar, S., ... Madabhushi, A. (2012, Apr). Concurrent segmentation of the prostate on MRI and CT via linked statistical shape models for radiotherapy planning. *Med Phys*, 39(4), 2214–2228. doi: 10.1118/1.3696376
- Cohen, I., Cohen, L. D., & Ayache, N. (1992). Using deformable surfaces to segment 3-D images and infer differential structures. *CVGIP: Image Understanding*, 56(2), 242–263. doi: 10.1016/1049-9660(92)90041-Z
- Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., & Marchal, G. (1995). Automated multi-modality image registration based on information theory. In *Information processing in medical imaging* (Vol. 3, pp. 264–274).
- Collins, D., Neelin, P., Peters, T., & Evans, A. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized talairach space. *Journal of computer assisted tomography*, 18(2), 192–205.
- Cootes, T. F., Edwards, G. J., & Taylor, C. J. (1998). Active appearance models. In *Computer vision – ECCV’98* (Vol. 1407, pp. 484–498). Springer Berlin / Heidelberg. doi: 10.1007/BFb0054760
- Cootes, T. F., & Taylor, C. J. (2004, March). *Statistical models of appearance for computer vision* (Tech. Rep.). University of Manchester.

- Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1992). Training models of shape from sets of examples. In *Proceedings of the british machine vision conference* (pp. 9–18). Springer. doi: 10.1.1.141.4020
- Danish Health and Medicines Authority. (2009, November). *Kræftprofil: Prostatakræft 2000-2007* (Tech. Rep.). Danish Health and Medicines Authority.
- Dearnaley, D. P., Khoo, V. S., Norman, A. R., Meyer, L., Nahum, A., Tait, D., ... Horwich, A. (1999). Comparison of radiation side-effects of conformal and conventional radiotherapy in prostate cancer: a randomised trial. *The Lancet*, 353(9149), 267–272. doi: 10.1016/S0140-6736(98)05180-0
- Deyo, R. A., Gray, D. T., Kreuter, W., Mirza, S., & Martin, B. I. (2005, June). United States trends in lumbar fusion surgery for degenerative conditions. *Spine*, 30(12), 1441–1445. doi: 10.1097/01.brs.0000166503.37969.8a
- Deyo, R. A., & Mirza, S. K. (2006, February). Trends and variations in the use of spine surgery. *Clinical Orthopaedics and Related Research*, 443, 139–146. doi: 10.1097/01.blo.0000198726.62514.75
- Deyo, R. A., Nachemson, A., & Mirza, S. K. (2004). Spinal-fusion surgery – the case for restraint. *New England Journal of Medicine*, 350(7), 722–726. doi: 10.1056/NEJMs031771
- Dryden, I. L., & Mardia, K. V. (1998). *Statistical shape analysis*. John Wiley & Sons.
- Erasmus, L., Hurter, D., Naudé, M., Kritzinger, H., & Acho, S. (2004). A short overview of MRI artefacts. *South African Journal of Radiology*, 8(2).
- Eskildsen, S., & Østergaard, L. (2006). Active surface approach for extraction of the human cerebral cortex from MRI. In R. Larsen, M. Nielsen, & J. Sporring (Eds.), *Medical image computing and computer-assisted intervention – MICCAI 2006* (Vol. 4191, pp. 823–830). Springer Berlin / Heidelberg. doi: 10.1007/11866763_101
- Feuerman, M., & Miller, A. R. (2008). Relationships between statistical measures of agreement: sensitivity, specificity and kappa. *Journal of Evaluation in Clinical Practice*, 14(5), 930–933. doi: 10.1111/j.1365-2753.2008.00984.x
- Flores-Tapia, D., Thomas, G., Venugopal, N., McCurdy, B., & Pistorius, S. (2008, August). Semi automatic MRI prostate segmentation based on wavelet multiscale products. In *Engineering in*

- medicine and biology society, 2008. EMBS 2008. 30th annual international conference of the IEEE* (pp. 3020–3023). doi: 10.1109/IEMBS.2008.4649839
- Fokdal, L. U., & Høyer, M. (2005, September). Senbivirkninger efter kurativ strålebehandling for cancer prostatae. *Ugeskrift for Læger, 167*(37), 3502–3507.
- Fu, T.-S., Chen, L.-H., Wong, C.-B., Lai, P.-L., Tsai, T.-T., Niu, C.-C., & Chen, W.-J. (2004). Computer-assisted fluoroscopic navigation of pedicle screw insertion an in vivo feasibility study. *Acta Orthopaedica, 75*(6), 730–735. doi: 10.1080/00016470410004102
- Ghose, S., Oliver, A., Martí and, R., Lladó and, X., Freixenet, J., Vilanova, J., & Meriaudeau, F. (2010, December). Prostate segmentation with texture enhanced active appearance model. In *Signal-image technology and internet-based systems (SITIS), 2010 sixth international conference on* (pp. 18–22). doi: 10.1109/SITIS.2010.14
- Gonzalez, R. C., & Woods, R. E. (2008). *Digital image processing* (third ed.). Pearson/Prentice Hall.
- Gray, H. (1918). *Gray's anatomy plates*. Retrieved from http://commons.wikimedia.org/wiki/Gray's_Anatomy_plates
- Hosten, N., & Liebig, T. (2002). *CT of the head and spine*. Thieme.
- Hu, S., & Collins, D. L. (2007). Joint level-set shape modeling and appearance modeling for brain structure segmentation. *NeuroImage, 36*(3), 672–683. doi: 10.1016/j.neuroimage.2006.12.048
- Huttenlocher, D., Klanderman, G., & Rucklidge, W. (1993, September). Comparing images using the hausdorff distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 15*(9), 850–863. doi: 10.1109/34.232073
- Ito, Y., Sugimoto, Y., Tomioka, M., Hasegawa, Y., Nakago, K., & Yagata, Y. (2008). Clinical accuracy of 3D fluoroscopy – assisted cervical pedicle screw insertion. *Journal of Neurosurgery, 9*(5), 450–453. doi: 10.3171/SPI.2008.9.11.450
- Jakobsen, H., Iversen, P., Mikines, K. J., Nørgaard, N., & Høyer, M. (2007, May). Behandling af lokaliseret prostatacancer. *Ugeskrift for Læger, 169*(20), 1902–1904.
- Jensen, C. M., Hermann, N., de Neergaard, L., Sørensen, A. J., Krausing-Vinther, J., Hartling, S., ... Arndal, J. (2009, May). *Bilag til årsrapport for arbejdet i task force vedr. strålebehandling 2008* (Tech. Rep.). Danish Health and Medicines Authority.

- Jønler, M., & Pedersen, K. V. (2007, May). Diagnose, udredning og opfølgning af patienter med prostatacancer. *Ugeskrift for Læger*, 169(20), 1889–1891.
- Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, 321–331.
- Kaus, M., Pekar, V., Lorenz, C., Truyen, R., Lobregt, S., & Weese, J. (2003, August). Automated 3-D PDM construction from segmented images using deformable models. *Medical Imaging, IEEE Transactions on*, 22(8), 1005–1013. doi: 10.1109/TMI.2003.815864
- Kiefer, J. (1953). Sequential minimax search for a maximum. *Proceedings of the American Mathematical Society*, 4(3), 502–506. doi: 10.2307/2032161
- Kim, Y., & Kim, D. (2009). A fully automatic vertebra segmentation method using 3D deformable fences. *Computerized Medical Imaging and Graphics*, 33(5), 343–352. doi: 10.1016/j.compmedimag.2009.02.006
- Kirby, M., & Sirovich, L. (1990, January). Application of the karhunen-loeve procedure for the characterization of human faces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(1), 103–108. doi: 10.1109/34.41390
- Kirschner, M., Becker, M., & Wesarg, S. (2011). 3D active shape model segmentation with nonlinear shape priors. In G. Fichtinger, A. Martel, & T. Peters (Eds.), *Medical image computing and computer-assisted intervention – MICCAI 2011* (Vol. 6892, pp. 492–499). Springer Berlin / Heidelberg. doi: 10.1007/978-3-642-23629-7_60
- Klein, S., van der Heide, U. A., Lips, I. M., van Vulpen, M., Staring, M., & Pluim, J. P. W. (2008). Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. *Medical Physics*, 35(4), 1407–1417. doi: 10.1118/1.2842076
- Klinder, T., Ostermann, J., Ehm, M., Franz, A., Kneser, R., & Lorenz, C. (2009). Automated model-based vertebra detection, identification, and segmentation in CT images. *Medical Image Analysis*, 13(3), 471–482. doi: 10.1016/j.media.2009.02.004
- Klinder, T., Wolz, R., Lorenz, C., Franz, A., & Ostermann, J. (2008). Spine segmentation using articulated shape models. In *Medical image computing and computer-assisted intervention – MICCAI 2008* (Vol. 5241, pp. 227–234). Springer Berlin / Heidelberg. doi: 10.1007/978-3-540-85988-8_28
- Koller, H., Acosta, F., Tauber, M., Fox, M., Martin, H., Forstner, R., ... Kässmann, H. (2008). Cervical anterior transpedicular screw fixation (ATPS) – part II. accuracy of manual insertion and pull-out

- strength of ATPS. *European Spine Journal*, 17(4), 539–555. doi: 10.1007/s00586-007-0573-x
- Kotani, Y., Abumi, K., Ito, M., Takahata, M., Sudo, H., Ohshima, S., & Minami, A. (2007). Accuracy analysis of pedicle screw placement in posterior scoliosis surgery: Comparison between conventional fluoroscopic and computer-assisted technique. *Spine*, 32(14), 1543–1550. doi: 10.1097/BRS.0b013e318068661e
- Kotil, K., & Bilge, T. (2008). Accuracy of pedicle and mass screw placement in the spine without using fluoroscopy: a prospective clinical study. *The Spine Journal*, 8(4), 591–596. doi: 10.1016/j.spinee.2007.04.002
- Laine, T., Lund, T., Ylikoski, M., Lohikoski, J., & Schlenzka, D. (2000). Accuracy of pedicle screw insertion with and without computer assistance: a randomised controlled clinical study in 100 consecutive patients. *European Spine Journal*, 9(3), 235–240. doi: 10.1007/s005860000146
- Lay, D. C. (2006). *Linear algebra and its applications* (Third ed.). Pearson.
- Leventon, M., Grimson, W., & Faugeras, O. (2000). Statistical shape influence in geodesic active contours. In *Computer vision and pattern recognition, 2000. proceedings. IEEE conference on* (Vol. 1, pp. 316–323). doi: 10.1109/CVPR.2000.855835
- Lorenz, C., & Krahnstover, N. (1999). 3D vertebrae segmentation using graph cuts with shape prior constraints statistical shape models for medical image segmentation. In *3-D digital imaging and modeling, second international conference on* (pp. 414–423). IEEE. doi: 10.1109/IM.1999.805372
- Ludwig, S. C., Kramer, D. L., Balderston, R. A., Vaccaro, A. R., Foley, K. F., & Albert, T. J. (2000, July). Placement of pedicle screws in the human cadaveric cervical spine: Comparative accuracy of three techniques. *Spine*, 25(13), 1655–1667.
- Martini, F. H. (2006). *Fundamentals of anatomy & physiology* (seventh ed.). Pearson Education.
- Mastmeyer, A., Engelke, K., Fuchs, C., & Kalender, W. A. (2006). A hierarchical 3D segmentation method and the definition of vertebral body coordinate systems for QCT of the lumbar spine. *Medical Image Analysis*, 10(4), 560–577. (Special Issue on Functional Imaging and Modelling of the Heart (FIMH 2005)) doi: 10.1016/j.media.2006.05.005

- Mazonakis, M., Damilakis, J., Varveris, H., Prassopoulos, P., & Gourtsoyiannis, N. (2001). Image segmentation in treatment planning for prostate cancer using the region growing technique. *British Journal of Radiology*, 74(879), 243–249.
- McConnell Brain Imaging Center. (2012, April). *Minc tool kit*. Retrieved from <http://www.bic.mni.mcgill.ca/ServicesSoftware/ServicesSoftwareMincToolKit>
- Mercier, L., Langø, T., Lindseth, F., & Collins, L. D. (2005). A review of calibration techniques for freehand 3-D ultrasound systems. *Ultrasound in Medicine & Biology*, 31(2), 143–165. doi: 10.1016/j.ultrasmedbio.2004.11.001
- Mercier, L., Maestro, R. F. D., Petrecca, K., Kochanowska, A., Drouin, S., Yan, C. X. B., ... Collins, D. L. (2011). New prototype neuronavigation system based on preoperative imaging and intraoperative freehand ultrasound: system description and validation. *International Journal of Computer Assisted Radiology and Surgery*, 6(4), 507–522. doi: 10.1007/s11548-010-0535-3
- Murase, H., & Nayar, S. (1995). Visual learning and recognition of 3-D objects from appearance. *International journal of computer vision*, 14(1), 5–24.
- Neo, M., Sakamoto, T., Fujibayashi, S., & Nakamura, T. (2005). The clinical risk of vertebral artery injury from cervical pedicle screws inserted in degenerative vertebrae. *Spine*, 30(24), 2800–2805.
- Østergaard, L., Eskildsen, S., Nielsen, J., Holmberg, A., Larsen, E., Fabrin, K., ... Carl, J. (2010). The use of a prostate stent for MR-CT registration in image-guided radiotherapy of prostate cancer. *Radiation Therapy & Oncology*, 96(Suppl. 1), S 417–S 418, No. 1238.
- Pasquier, D., Lacornerie, T., Vermandel, M., Rousseau, J., Lartigau, E., & Betrouni, N. (2007). Automatic segmentation of pelvic structures from magnetic resonance images for prostate cancer radiotherapy. *International Journal of Radiation Oncology*Biophysics*, 68(2), 592–600. doi: 10.1016/j.ijrobp.2007.02.005
- Rajasekaran, S., Vidyadhara, S., Ramesh, P., & Shetty, A. P. (2007, January). Randomized clinical study to compare the accuracy of navigated and non-navigated thoracic pedicle screws in deformity correction surgeries. *Spine*, 32(2), E56–E64. doi: 10.1097/01.brs.0000252094.64857.ab
- Rasmussen, S., Jensen, C. M., Iversen, M. G., & Kehlet, H. (2009, September). Lumbal spondylodese for degenerativ ryglidelse i Danmark i 2005–2006. *Ugeskrift for Læger*, 171(39), 2804–2807.

- Roberts, M., Cootes, T., Pacheco, E., Oh, T., & Adams, J. (2009). Segmentation of lumbar vertebrae using part-based graphs and active appearance models. In G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, & C. Taylor (Eds.), *Medical image computing and computer-assisted intervention – MICCAI 2009* (Vol. 5762, pp. 1017–1024). Springer Berlin / Heidelberg. doi: 10.1007/978-3-642-04271-3_123
- Rockafellar, R. T., & Wets, R. J.-B. (1998). *Variational analysis*. Springer.
- Shen, H., Litvin, A., & Alvino, C. (2008). Localized priors for the precise segmentation of individual vertebrae from CT volume data. In *Medical image computing and computer-assisted intervention – MICCAI 2008* (Vol. 5241, pp. 367–375). Springer Berlin / Heidelberg. doi: 10.1007/978-3-540-85988-8_44
- Stegmann, M. B., & Gomez, D. D. (2002, March). *A brief introduction to statistical shape analysis* (Tech. Rep.). Informatics and Mathematical Modelling, Technical University of Denmark, DTU. Retrieved from http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=403
- Tsai, A., Wells, W., Tempany, C., Grimson, E., & Willsky, A. (2004). Mutual information in coupled multi-shape model for medical image segmentation. *Medical Image Analysis*, 8(4), 429–445. doi: 10.1016/j.media.2004.01.003
- Tsai, A., Yezzi, J., A., Wells, I., W., Tempany, C., Tucker, D., Fan, A., ... Willsky, A. (2001). Model-based curve evolution technique for image segmentation. In *Computer vision and pattern recognition, 2001. CVPR 2001. proceedings of the 2001 IEEE computer society conference on* (Vol. 1, pp. I-463–I-468). doi: 10.1109/CVPR.2001.990511
- Tsai, A., Yezzi, J., A., Wells, W., Tempany, C., Tucker, D., Fan, A., ... Willsky, A. (2003, February). A shape-based approach to the segmentation of medical imagery using level sets. *Medical Imaging, IEEE Transactions on*, 22(2), 137–154. doi: 10.1109/TMI.2002.808355
- Turk, M., & Pentland, A. (1991, 2011/08/16). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 71–86. doi: 10.1162/jocn.1991.3.1.71
- Villeirs, G. M., & Meerleer, G. O. D. (2007). Magnetic resonance imaging (MRI) anatomy of the prostate and application of MRI in radiotherapy planning. *European Journal of Radiology*, 63(3), 361–368. doi: 10.1016/j.ejrad.2007.06.030

- Vordermark, D., Schwab, M., Ness-Dourdoumas, R., Sailer, M., Flen-tje, M., & Koelbl, O. (2003). Association of anorectal dose–volume histograms and impaired fecal continence after 3D conformal ra-diotherapy for carcinoma of the prostate. *Radiotherapy and Oncol-ogy*, 69(2), 209–214. doi: 10.1016/j.radonc.2003.07.002
- Weese, J., Kaus, M., Lorenz, C., Lobregt, S., Truyen, R., & Pekar, V. (2001). Shape constrained deformable models for 3D medical im-age segmentation. In *Information processing in medical imag-ing* (Vol. 2082, pp. 380–387). Springer Berlin / Heidelberg. doi: 10.1007/3-540-45729-1_38
- Yan, C. X. B., Goulet, B., Pelletier, J., Chen, S. J.-S., Tampieri, D., & Collins, D. L. (2011). Towards accurate, robust and practi-cal ultrasound-CT registration of vertebrae for image-guided spine surgery. *International Journal of Computer Assisted Radiology and Surgery*, 6(4), 523–537. doi: 10.1007/s11548-010-0536-2
- Zijdenbos, A. P., Dawant, B. M., Margolin, R. A., & Palmer, A. C. (1994). Morphometric analysis of white matter lesions in MR im-ages: method and validation. *Medical Imaging, IEEE Transactions on*, 13(4), 716–724. doi: 10.1109/42.363096
- Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M., Kaus, M. R., Haker, S. J., ... Kikinis, R. (2004, Feb). Statistical validation of im-age segmentation quality based on a spatial overlap index. *Acad Radiol*, 11(2), 178–189.

Appendices

Representing 3-D pose

Three-dimensional pose is represented by translation in each direction, rotation around each axis, and scaling.

Translation is represented by the translation parameters $\vec{t} = [t_x, t_y, t_z]^T$. In homogeneous coordinates translation is applied by multiplying the point coordinates and the following matrix:

$$\mathbf{T}_t = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{A.1})$$

Rotation is represented by the rotation angles ϕ , θ , and ψ , representing rotation around the x , y and z axis respectively. In homogeneous coordinates rotation is applied by first translating the image to put the center of rotation in $(0, 0, 0)$, multiplying the coordinates with one or more of the following matrices, and finally translating back to the original location:

$$\mathbf{T}_\phi = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi & 0 \\ 0 & \sin\phi & \cos\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{A.2})$$

$$\mathbf{T}_\theta = \begin{bmatrix} \cos\theta & 0 & \sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{A.3})$$

$$\mathbf{T}_\psi = \begin{bmatrix} \cos\psi & -\sin\psi & 0 & 0 \\ \sin\psi & \cos\psi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{A.4})$$

The center of rotation is represented by $\vec{c} = [c_x, c_y, c_z]^T$, and the translation matrix is:

$$\mathbf{T}_c = \begin{bmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 1 & -c_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{A.5})$$

Scaling is applied by multiplying this matrix:

$$\mathbf{T}_s = \begin{bmatrix} e^s & 0 & 0 & 0 \\ 0 & e^s & 0 & 0 \\ 0 & 0 & e^s & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{A.6})$$

where s is the natural logarithm of the desired scaling. This makes $s = 0$ correspond to no change in scale.

By combining the transformation matrices and pose parameters, pose is represented by the pose vector \vec{t} . A set of coordinates \vec{x} is transformed by multiplying each transformation matrix in turn:

$$\vec{t} = [\phi, \theta, \psi, s, t_x, t_y, t_z]^\top \quad (\text{A.7})$$

$$\vec{x}' = \mathbf{T}_t \mathbf{T}_c^{-1} \mathbf{T}_s \mathbf{T}_\psi \mathbf{T}_\theta \mathbf{T}_\phi \mathbf{T}_c \vec{x} \quad (\text{A.8})$$

Note that first the center of rotation is moved to the origin, and then rotations are applied to rotate around this center. After rotations scaling is applied, which also depends on the center of rotation. Then the image is translated back to the original position, and finally the desired translation is applied.

This way of representing pose (especially scaling) ensures the following properties, which are important for the AAM predictions of parameter change [Cootes & Taylor, 2004]:

- The zero parameter vector must imply no change in pose:
 $\vec{t} = \mathbf{0} \implies \mathbf{T}(\vec{t}) = \mathbb{1}$.
- The pose change must satisfy (at least for small pose changes):
 $\mathbf{T}(\vec{t}_1) \mathbf{T}(\vec{t}_2) \approx \mathbf{T}(\vec{t}_1 + \vec{t}_2)$.

Pipeline implementation

The preprocessing and validation steps are implemented as PSOM pipelines. Each pipeline is composed of a series of bricks, and each brick takes files as inputs and produces files as outputs. The pipelining system handles dependencies between the bricks, and re-runs a brick if the inputs or options change. All child bricks will be re-run too, because their inputs have changed. [Bellec et al., 2012]

B.1 Pipelines

Pipelines can be combined into larger pipelines by routing the output of one pipeline to the input of another pipeline. The following pipelines are implemented:

DICOM preprocessing

This pipeline takes DICOM and DICOM-RT images as input, and produces a MINC format volume, binary segmentation, distance map, and MNI .obj representation of the manual segmentation (Fig. B.1). All outputs are produced both in the original coordinate system and in the coordinate system of the common reference. The pipeline also computes the parameter change prediction matrix for changes in pose and texture.

MINC preprocessing

This pipeline takes MINC images and segmentations as input, and produces the same output as the DICOM preprocessing pipeline (Fig. B.2).

Training

This pipeline trains the appearance model on the input. It takes pose parameter change prediction matrices, MINC volumes, and MINC distance maps as input, and computes an appearance model and full parameter change prediction matrix.

Validation

This pipeline runs leave-one-out cross-validation on the input. It takes the same input as the training pipeline, as well as the original image volumes, segmentations, distance maps and MNI .obj representations for validation (Fig. B.3). The pipeline produces a MATLAB mat-file with combined performance metrics for analysis.

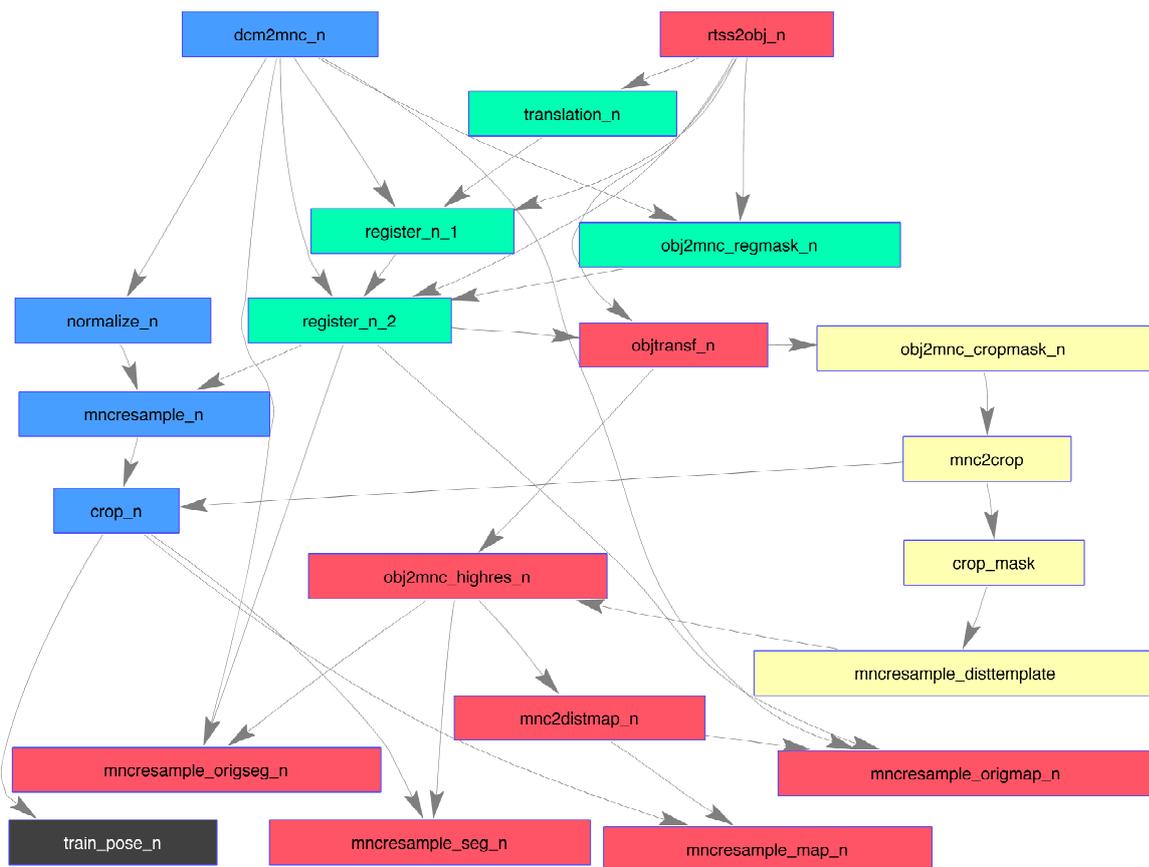


Figure B.1: The DICOM preprocessing pipeline. Bricks named with $_n$ are run for each input image, other bricks are run only once. Steps are color-coded according to: *Image preprocessing*, *Segmentation preprocessing*, *Registration*, and *Cropping*.

The training and validation pipelines can be combined with one of the preprocessing pipelines, depending on the file format of the original data.

B.2 Bricks

The pipelines utilize a number of commands, or bricks. Each brick runs independently from other bricks, and is only dependent on input files and options. A brick produces a set of output files, which can be used as input for subsequent bricks. Many bricks are wrappers for underlying MINC command-line tools, while other bricks are pure MATLAB operations. The pipelines use a combination of the following bricks:

DICOM to MINC (dcm2mnc)

Wrapper for dcm2mnc from the MINC tool kit. Converts a set of DICOM image slices to a single MINC volume.

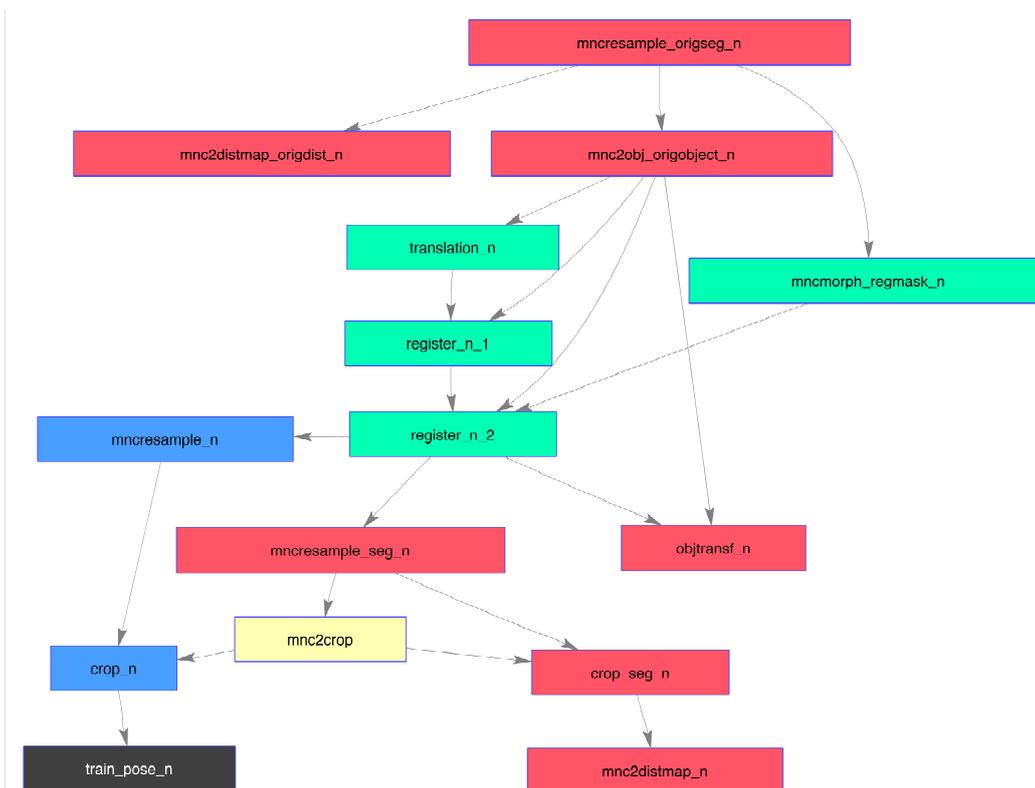


Figure B.2: The MINC preprocessing pipeline. Bricks named with $_n$ are run for each input image, other bricks are run only once. Steps are color-coded according to: *Image preprocessing*, *Segmentation preprocessing*, *Registration*, and *Cropping*.

DICOM-RT to MNI .obj (rtss2obj)

Convert structures in a DICOM-RT input file to MNI .obj representation. Each structure is saved as an individual object file. Optionally only structures with names matching a given regular expression are saved.

MINC to distance map (mnc2distmap)

Wrapper for `itk_distance`. Computes the signed distance map of the input volumes. The input volumes must have uniform resolution for the distance values to be meaningful, or the input can optionally be resampled to uniform resolution before computing the distance map. The distance map is resampled back into the resolution of the input image.

MINC to MNI .obj (mnc2obj)

Wrapper for `marching_cubes`. Computes the MNI .obj representation of the input image using marching cubes. The input should be a signed distance map or binary image. Inside voxels must have larger

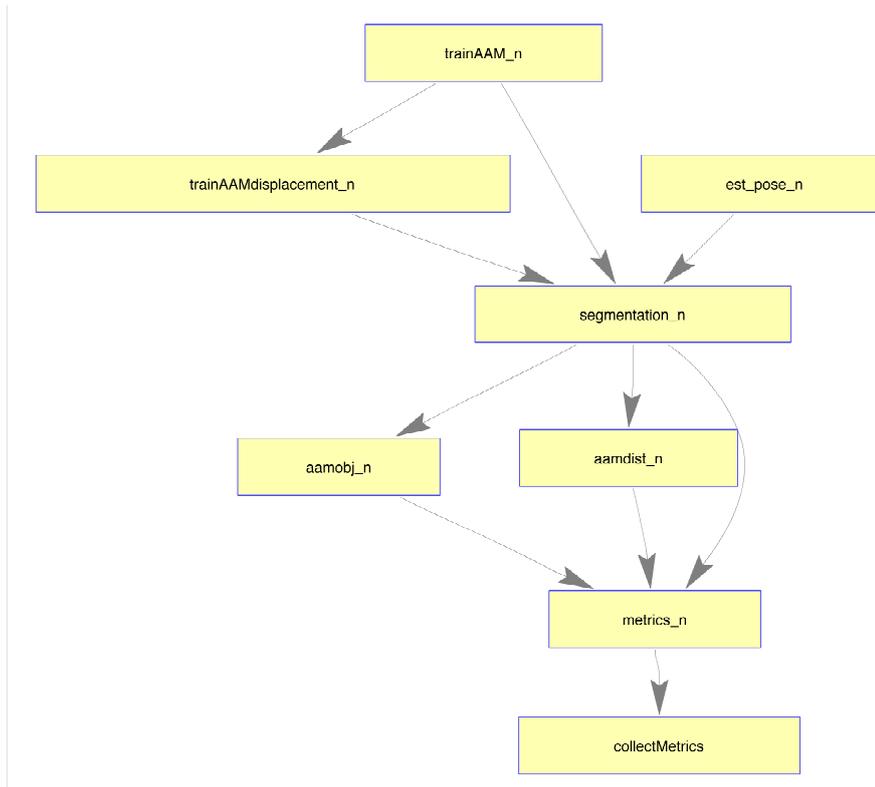


Figure B.3: The validation pipeline. Bricks ending in *_n* are run for each input image, the last brick is run only once. For each image an appearance model is trained on the other images and used for segmentation, before computing performance metrics.

values than outside voxels. To fulfill this constraint the input volume can optionally be negated before running marching cubes.

MNI .obj to MINC (obj2mnc)

Wrapper for `surface_mask2`. Computes a binary MINC mask with the same sampling as a reference volume. Optionally performs erosion or dilation of the binary mask (see below for details).

Morphology (mncmorph)

Performs erosion or dilation of a binary MINC volume using MATLAB `imerode` or `imdilate`. A 26-connected morphology kernel of a specified radius is used. The radius is in units of mm, and the actual size of the kernel depends on the voxel size.

MINC to crop mask (mnc2crop)

Merges a set of MINC masks into a single mask. The output is the union of the input masks.

Crop (crop)

Wrapper for `autocrop`. Crops an image volume to the rectangular bounds in a binary mask.

Volume registration (register)

Wrapper for `minctracc`. Supports all `minctracc` options, including center of translations and rotations, initial transformation guess, and volume masks. Outputs the transformation matrix transforming the input volume to the target.

Resample MINC volume (mncresample)

Wrapper for `mincresample` and `itk_resample`. Resamples the input MINC volume with specified settings. The volume can optionally be transformed using an input transformation matrix. The sampling options of a specified reference volume can be used, and the image can be resampled to uniform resolution. Interpolation is done using nearest neighbor (`mincresample`) or B-splines (`itk_resample`).

Transform MNI .obj file (objtransf)

Wrapper for `transform_objects`. Transforms an MNI .obj file using an input transformation matrix.

Grayscale normalization (normalize)

Normalizes an input MINC volume to zero mean and a specified standard deviation of grayscale intensity values.

Train appearance model (train_model)

Trains an appearance model on the input MINC volumes and distance maps. Outputs the appearance model as a MATLAB mat-file.

Train appearance parameter displacements (train_displace_model)

Systematically displaces each parameter in the input appearance model and computes the parameter change prediction matrix. Outputs the matrix as a mat-file.

Train pose parameter displacements (train_displace_pose)

Systematically displaces each pose parameter and computes the parameter change prediction matrix for pose parameters using the input MINC volume. Outputs the matrix as a mat-file.

Segmentation (segmentation)

Does segmentation of a MINC volume using a specified appearance model, parameter change prediction matrix, and initial parameter guess. Outputs the binary segmentation and computed distance map of the appearance model as MINC files with the same sampling options as the input volume.

Metrics (metrics)

Computes performance metrics by comparing two segmentations. See chapter 8 for details. Output is a MATLAB mat-file containing the metrics. Some metrics are computed using `volume_similarity`, others are computed using a combination of Matlab and other tools.

Collect metrics (collect_metrics)

Combines mat-files containing performance metrics into a single mat-file.