Generative Interpretation of Medical Images

Mikkel B. Stegmann

Informatics and Mathematical Modelling
Technical University of Denmark
Ph.D. Thesis No. 127
Kgs. Lyngby 2004

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kgs. Lyngby, Denmark
Phone +45 4525 3351, Fax +45 4588 2673
www.imm.dtu.dk

Thesis submitted 21st January and defended 1st April 2004.
Document version 20th April 2004.
Typeset using LaTeX 2_ϵ.
Printed by IMM, DTU.
1st edition.

IMM-PHD-2004-127
ISSN 0909-3192

Preface

This thesis has been prepared at the Image Analysis and Computer Graphics group at Informatics and Mathematical Modelling – IMM and submitted to the Technical University of Denmark – DTU, in partial fulfilment of the requirements for the degree of Doctor of Philosophy, Ph.D., in applied mathematics.

The work herein represents selected parts of the two years of research work allotted within the three years of the Danish Ph.D. study. The thesis consists of seven research papers and an introductory part containing an overview and giving some background information.

The topic treated is medical image analysis with applications to cardiac MRI, brain MRI and chest radiographs. Basic knowledge of image analysis, statistics, linear algebra and general mathematics are assumed. The work was carried out in collaboration with the Danish Research Centre for Magnetic Resonance – DRCMR, H:S Hvidovre Hospital (a part of Copenhagen University Hospital), and the Department of Diagnostic Imaging at St. Olavs Hospital, Trondheim University, Norway.

Research funding was provided by the Danish Medical Research Council, grant number 52-00-0767. The project was supervised by Bjarne K. Ersbøll (IMM), Rasmus Larsen (IMM) and Henrik B. W. Larsson (St. Olavs Hospital/DRCMR).

Kgs. Lyngby, January 2004

Mikkel B. Stegmann

Acknowledgements

Writing a doctoral thesis from start to end is a sizeable task. Luckily, quite a few have kindly offered their help and company during the past three years. This section expresses my gratitude.

However, without the funding from the Danish Medical Research Council, all of this wouldn't have taken place. I sincerely thank you for believing in this project, which has been the most fun, interesting and rewarding part of my schooling.

First of all I would like to thank my supervisors Rasmus Larsen, Bjarne K. Ersbøll and Henrik B. W. Larsson for your support, encouragement and discussions, in addition to your pleasant company throughout the years.

Many thanks to my co-authors for making this thesis more than the work of a one-man-band. I have truly enjoyed and learned from all our discussions. In order of appearance: Bjarne K. Ersbøll, Rasmus Larsen, Søren Forchhammer, Timothy F. Cootes, Rhodri H. Davies, Charlotte Ryberg, Bram van Ginneken, Marco Loog, Hildur Ólafsdóttir and Henrik B. W. Larsson.

I'm indebted to Dorthe Pedersen, Bjørn A. Grønning, Jens Chr. Nilsson, Lars G. Hanson, Egill Rostrup, Charlotte Ryberg and Torben Lund from the Danish Research Centre for Magnetic Resonance (DRCMR) for supplying me with MR data and letting me in on all the intricate details of MR scanning, cardiac anatomy, et cetera.

I owe a great debt to Tim Cootes and Chris Taylor for letting me visit the Victoria University of Manchester during my external research stay. Tim, Chris, Christine, Gavin, Kola, Dave and all you other folks at ISBE, you made it a great stay in every sense. Thanks.

Thanks are due to all past and present members of the Image Analysis and Computer Graphics group at IMM for an always rewarding and pleasant atmosphere. Rune Fisker deserves special thanks for getting me started in this field. Besides

taking part in several enjoyable conference trips Rasmus Reinhold Paulsen also reviewed this thesis in great detail. Thanks. Former officemates Klaus Baggesen Hilger and Lars Pedersen have in many ways been essential to my education in being a Ph.D. student. In particular, by introducing me to the art of conference participation. They even reviewed this thesis long after obtaining their own degrees. Thank you for all the fruitful discussions, collaborations and your great company.

Karlheinz Brandenburg, Justin Frankel, Nick Foster and Mike Lai are thanked for their great efforts in supplying infrastructure and content pertinent to the creation of this thesis.

Finally, heartfelt thanks go to Katharina and Nikoline for all your love and support. Even during the stressful time of a thesis write-up. Thanks.

Abstract

This thesis describes, proposes and evaluates methods for automated analysis and quantification of medical images. A common theme is the usage of generative methods, which draw inference from unknown images by synthesising new images having shape, pose and appearance similar to the analysed images. The theoretical framework for fulfilling these goals is based on the class of Active Appearance Models, which has been explored and extended in case studies involving cardiac and brain magnetic resonance images (MRI), and chest radiographs.

Topics treated include model truncation, compression using wavelets, handling of non-Gaussian variation by means of cluster analysis, correction of respiratory noise in cardiac MRI, and the extensions to multi-slice two-dimensional time-series and bi-temporal three-dimensional models.

The medical applications include automated estimation of: left ventricular ejection fraction from 4D cardiac cine MRI, myocardial perfusion in bolus passage cardiac perfusion MRI, corpus callosum shape and area in mid-sagittal brain MRI, and finally, lung, heart, clavicle location and cardiothoracic ratio in anterior-posterior chest radiographs.

Resumé

Denne afhandling beskriver, udvikler og evaluerer metoder til automatiseret analyse og opmåling af medicinske billeder. Et fælles tema er anvendelsen af såkaldte *generative* metoder, der muliggør estimering af latent information fra ukendte billeder ved hjælp af processer, der er i stand til at generere syntetiske billeder, med samme form, positur og udseende som de analyserede billeder. Metodegrundlaget for dette arbejde er modelklassen Active Appearance Models, som er blevet udforsket og udvidet i studier af magnetisk resonansbilleder (MRI) af hjerte og hjerne, samt røntgenbilleder af brystkassen (thorax).

I det rapporterede arbejde er en lang række delproblemer behandlet. Disse inkluderer: model beskæring, kompression via wavelets, håndtering af ikke-Gaussisk variation via klyngemodellering, korrektion af respiratorisk støj i hjerte MRI, samt udvidelser til todimensionale multiskive-tidsserier og bi-temporale tredimensionale modeller.

Medicinske anvendelser af de opnåede resultater inkluderer estimering af: venstre ventrikels uddrivningsfraktion (*ejection fraction*) fra hjerte MRI i fire dimensioner, blodgennemstrømning i hjertemuskelen fra perfusions MRI, form og areal af hjernebjælken (*corpus callosum*) fra mid-sagittal hjerne MRI, samt cardiothoracic ratio og position af lunge, hjerte og kraveben fra thoraxrøntgen.

Contents

Part I

Generative Interpretation of Medical Images

CHAPTER 1

Introduction

The health sector in Europe, USA and other parts of the world faces tremendous challenges due to an aging population and an ever-increasing advent of new treatments. This, in combination with an increasing incidence of age and life style related diseases, has lead to a vastly increased demand for treatment and monitoring. With conflicting political demands of limited or even no increase in budgets, this combines to a grievous if not untenable situation. This must be met with increased productivity. A recognised and all-important instrument for increasing productivity is correct and early diagnosis.

Among the most important diagnostic tools is *medical imaging*. This is a group of non-invasive techniques – pioneered by Wilhelm Roentgen – for visual probing of the human body. Today these are even capable of producing detailed volumetric images of dynamic processes, such as the beating heart.

While the impressive range of sophisticated and versatile medical imaging devices – among other things – supply the world with dazzling acronyms such as CT/DXA/MR/PET/SPECT/US, they also accentuate the need for a shift from manually assessed images towards efficient, accurate and reproducible computer-based methods. These methods should aim at assisting medical experts in their decisions by providing them with quantitative measures inferred from the above-mentioned imaging modalities.

This thesis deals with computerised interpretation of medical images. In detail it seeks to explore and develop methods that can answer questions such as:

- Where is the heart located in this image?
- What is the stroke volume of this heart?
- What is the cross-sectional area of a set of nerve fibres in the brain?
- Is ratio between lung and heart abnormal in this patient?

Figure 1.1: Image interpretation using a priori knowledge. What is depicted here? Courtesy of Preece et al. [181].

Tip: Try looking for a Dalmatian dog sniffing leaves in a park.

Further, such methods should also provide a sound basis for answering more abstract questions akin to:

- Is the blood supply reduced in this heart?
- If so, where?
- Does the shape of this brain structure differ significantly from those of normal subjects?

The ability to answer questions in line with the above is what this thesis regards as *interpretation*. This pertains not only to being able to partition an image into object and background, but also to provide descriptions of functional properties and relations inferred from image data. The aim of providing such high-level descriptions can also be denoted 'image understanding', a process for which a solution seldom can be obtained by a divide-and-conquer strategy. Consider Figure 1.1. A bottom-up approach that detects local features first and eventually seeks to combine these into a description of the subject depicted is indeed likely to fail.

The constructivist theorists within cognitive psychology believe that the process of seeing is an active process in which our world is constructed from both the retinal view and prior knowledge [181]. Without a priori knowledge, it would never have been possible to decipher the black blobs of Figure 1.1. This is the main assumption behind the constructivist approach [181]. Namely, that visual perception involves the intervention of representations and memories such as "dog" and "park". Mundy [167] also stresses this point (pp. 1213, l. 5–8):

> *"...This process of* recognition*, literally to RE-cognize, permits an aggregation of experience and the evolution of relationships between objects based on a series of observations."*

These thoughts are essential to the motivation for – and design and usage of – knowledge-based models in medical image analysis. They also form a natural core philosophy for this thesis: To capture human prior knowledge about anatomy, tissue appearance, modality characteristics, et cetera and efficiently employ this to infer functional information from new data. In our case though, the retinal view mentioned above is substituted with a likelihood derived from MR scanners and X-ray devices.

Specifically, this thesis focuses on one particular class of models, which provides a concrete implementation of model-based image interpretation that utilises strong priors derived from training data. This class was introduced by Edwards et al. [88] and Cootes et al. [52] in 1998 and denoted Active Appearance Models (AAMs).

AAMs are highly flexible models trained on annotated image data. This training data should consist of representative solutions to the targeted problem. An example could be localisation of the heart muscle. These models can then subsequently be employed to answer questions similar to the ones stated above. The basic principle for doing so is to match a model to an unknown image and thereby inferring knowledge based on the converged model configuration. This process is based on the *generative* property of AAMs; namely that these models are capable of synthesising near photo-realistic images of the object class they model. As such model-to-image matching is carried out by a pixel-wise matching.

Hence, the thesis title *Generative Interpretation of Medical Images*.

In the following chapters, AAMs are explored and extended with applications to cardiac and brain magnetic resonance imaging (MRI), chest radiographs and face images. Notice that when the term *unsupervised* is used, it refers to the *usage* of AAMs, contrary to the building process in AAMs, which is highly *supervised*.[1]

Finally, it should be mentioned that projects within medical image analysis are typically highly interdisciplinary. This thesis is no exception. It will draw on re-

[1] Notice that the paper in Chapter 10 relates *supervised* to the training process.

sults from multivariate statistics, numerical analysis, linear algebra, wavelet theory, medicine, MR physics, computational geometry, computer science, computer graphics, et cetera. Any presentation spanning such a broad spectrum will inevitably lack depth in some of the domains touched upon. The treatment of particular topics may therefore appear too superficial to some. Insofar this is possible; it is sought redeemed by providing pointers to relevant literature.

1.1 Objectives

A primary goal of this thesis work has been to construct *complete systems* that allows for inference of functional indices without manual interaction and preferably within a reasonable timeframe. Various subtopics have been treated in detail but not to an extent where they compromised the primary goal.

In addition, much effort has been put into fulfilling the secondary goal of providing *quantitative* analyses of the presented methods. Surprisingly, this seems not always to be the case in the medical image analysis literature. Unfortunate, since methods presenting qualitative results in a few subjects and completely lacking quantitative validation are of very little use to the clinical practice, which should remain the ultimate goal.

That being said, the available number of subjects for the studies presented here, does typically not qualify for a sufficient clinical validation. However, methods are presented along with schemes for quantitative validation that can be readily used in future validation studies; either being prospective or retrospective.

A final objective has been to extend the typical dissemination of knowledge by adding *transparency* to the research insofar possible. At the heart of this lies an ambition about letting algorithms and reference data sets being publicly available along with corresponding reference performance measures. In that way new and existing methods can be compared on a fair and transparent basis by eliminating, or at least significantly alleviating, the tedious process of re-implementing competing methods and re-generating similar training data (if at all possible), et cetera. Further, this should enable understanding of intricate details – or subtleties – otherwise obfuscated or omitted in research papers (typically having a fixed length requirement).

1.2 Thesis Overview

The thesis is comprised of two main parts; an overview and a background part (in which this section is placed) and a part consisting of seven research papers demonstrating a selection of the scholarly work carried out. Papers are organised by increasing data complexity going from 2D images to 2D time-series and ending with 4D data in the form of 3D time-series. Brief introductions to the chapters in Part II

are given in the end of this section. As each of these are self-contained *per se*, apologies are given for the inevitable overlaps that occur.

Reading the thesis in full should be carried out by reading Part II just before proceeding to Chapter 5.

Co-authors of Part II and their affiliations are listed in order of apperance below.

- Bjarne K. Ersbøll
  Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Denmark
- Rasmus Larsen
  Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Denmark
- Søren Forchhammer
  Research Centre COM, Technical University of Denmark, DTU, Denmark
- Timothy F. Cootes
  Division of Imaging Science and Biomedical Engineering, University of Manchester, England
- Rhodri H. Davies
  Centre for Neuroscience, Howard Florey Institute, University of Melbourne, Australia
- Charlotte Ryberg
  Danish Research Centre for Magnetic Resonance, H:S Hvidovre Hospital, Denmark
- Bram van Ginneken
  Image Sciences Institute, University Medical Center Utrecht, The Netherlands
- Marco Loog
  Image Sciences Institute, University Medical Center Utrecht, The Netherlands
- Hildur Ólafsdóttir
  Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Denmark
- Henrik B. W. Larsson
  Department of Diagnostic Imaging, St. Olavs Hospital, Trondheim University, Norway

Chapter 7 introduces the core concepts alongside the typical mathematical notation used in AAMs. In particular, the process of calculating model parameter updates when fitting AAMs to unseen images is described and two rivalling methods are presented and compared empirically. The issue of model truncation by means of parallel analysis is treated and evaluated. This chapter also presents a case study of face images and cardiac MRI richly illustrating the various AAM parts. Further, a noise-sensitivity study on cardiac MRI is given. An additional overall purpose is to describe the publicly available 2D AAM implementation, which produced the results of this chapter.

Chapter 8 deals with compression of the typically sizeable texture models in AAMs. The orthogonal Haar wavelet and the bi-orthogonal CDF 9-7 wavelets are evaluated on a cohort of face images at various compression ratios. Further, a simple weighting scheme is presented, which allows for substantial compression with a *simultaneous* improvement in registration accuracy.

Chapter 9 describes how the corpus callosum brain structure can be automatically and rapidly located and analysed in MRI using AAMs employing minimum description length (MDL) shape modelling. Further, an extended coarse-to-fine scheme is presented along with an assessment of the importance of establishing proper texture sampling regions.

Chapter 10 comprises a thorough comparative study of three fully automated methods for segmentation of posterior-anterior chest radiographs. All three are generic methods that require training data, which in this case consisted of annotated lung fields, clavicles and heart in images from a publicly available database. These annotations have also been made publicly available to facilitate future comparative studies. The three methods compared are: Active Shape Models, Active Appearance Models, and a pixel classification scheme. The performance of these and a second human observer are compared against a gold standard annotation using a measure of shape overlap and the cardiothoracic ratio (CTR).

Chapter 11 presents a novel method for registration of single and multi-slice cardiac perfusion MRI time-series obtained from patients with acute myocardial infarction. Shape models are based on correspondences obtained by MDL. Texture modelling is hampered by the pronounced non-Gaussian appearance of perfusion MRI. A solution based on an ensemble of linear texture models obtained from a unsupervised classification is described. Further, prior models on shape and pose, which exploit inherent properties of perfusion MRI time-series are proposed and evaluated.

Chapter 12 discusses the problems in obtaining 'truly' three-dimensional temporal data from slice time-series of cardiac MRI acquired using breath-hold. A rapid method that runs without supervision and approximates a solution to this ill-posed problem, is presented and evaluated in four-dimensional cine MRI from obese normals.

Chapter 13 treats the problem of automatically obtaining estimates of global cardiac function from four-dimensional cardiac MRI. An unsupervised bi-temporal three-dimensional AAM is proposed and evaluated. This AAM models

the end-diastole and end-systole of the cardiac cycle and is as such capable of estimating the ejection fraction (EF), the left ventricular mass (LVM), et cetera.

1.3 Publications

In addition to this thesis, results have been reported by means of journal papers, conference papers, conference abstracts, technical reports, et cetera. This section enumerates these contributions ordered by their relations to the chapters in Part II. The first entry, shown in bold face, denotes the publication comprising the chapter.

Chapter 7

[213] M. B. Stegmann, B. K. Ersbøll, and R. Larsen. FAME – a flexible appearance modelling environment. *IEEE Trans. on Medical Imaging*, 22(10):1319–1331, 2003.

[206] M. B. Stegmann. Analysis and segmentation of face images using point annotations and linear subspace techniques. Technical Report IMM-REP-2002-22, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, aug 2002.

[209] M. B. Stegmann. The AAM-API: An open source active appearance model implementation. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003, 6th Int. Conference, Montréal, Canada*, LNCS 2879, pages 951–952. Springer, nov 2003.

Chapter 8

[218] M. B. Stegmann, S. Forchhammer, and T. F. Cootes. Wavelet enhanced appearance modelling. In *International Symposium on Medical Imaging 2004, San Diego CA, SPIE*. SPIE, 2004 (in press).

[217] M. B. Stegmann and S. Forchhammer. On exploiting wavelet bases in statistical region-based segmentation. In *Proc. 11th Danish Conference on Pattern Recognition and Image Analysis*, volume 1, pages 75–82, Copenhagen, Denmark, aug 2002. DIKU.

[216] M. B. Stegmann and S. Forchhammer. On decomposing object appearance using PCA and wavelet bases with applications to image segmentation. In Hans Joachim Werner, editor, *MATRIX'02, Eleventh International Workshop on Matrices and Statistics*, page 21. Informatics and Mathematical Modelling, Technical University of Denmark, DTU, sep 2002.

Chapter 9

[212] M. B. Stegmann, R. H. Davies, and C. Ryberg. Corpus callosum analysis using MDL-based sequential models of shape and appearance. In *International Symposium on Medical Imaging 2004, San Diego CA, SPIE*. SPIE, feb 2004 (in press).

[211] M. B. Stegmann and R. H. Davies. Automated analysis of corpora callosa. Technical Report IMM-REP-2003-02, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, mar 2003.

Chapter 10

[245] B. van Ginneken, M. B. Stegmann, and M. Loog. Segmentation of anatomical structures in chest radiographs using supervised methods: A comparative study on a public database. *Medical Image Analysis*, 2004 (submitted).

Chapter 11

[227] M. B. Stegmann, H. Ólafsdóttir, and H. B. W. Larsson. Unsupervised motion-compensation of multi-slice cardiac perfusion MRI. *Invited contribution for the FIMH special issue in Medical Image Analysis*, 2004 (submitted).

[224] M. B. Stegmann and H. B. W. Larsson. Motion-compensation of cardiac perfusion MRI using a statistical texture ensemble. In *Functional Imaging and Modeling of the Heart, FIMH 2003*, volume 2674 of *LNCS*, pages 151–161, Lyon, France, 2003. Springer.

[223] M. B. Stegmann and H. B. W. Larsson. Fast registration of cardiac perfusion MRI. In *Proc. International Society of Magnetic Resonance In Medicine – ISMRM 2003, Toronto, Ontario, Canada*, page 702, Berkeley, CA, USA, 2003. ISMRM.

Chapter 12

[225] M. B. Stegmann and H. B. W. Larsson. Rapid and unsupervised correction of respiratory-induced motion in 4D cardiac cine MRI. 2004 (to be submitted).

[222] M. B. Stegmann, R. Larsen, and H. B. W. Larsson. Unsupervised correction of physiologically-induced slice-offsets in 4D cardiac MRI. *Journal of Cardiovascular Magnetic Resonance (7th Annual SCMR Meeting, Barcelona, Spain)*, 6(1):451–452, 2004.

[226] M. B. Stegmann, J. C. Nilsson, and B. A. Grønning. Automated segmentation of cardiac magnetic resonance images. In *Proc. International Society of Magnetic Resonance In Medicine – ISMRM 2001*, volume 2, page 827. ISMRM, 2001.

Chapter 13

[210] M. B. Stegmann. Bi-temporal 3D active appearance modelling with applications to unsupervised ejection fraction estimation from 4D cardiac MRI. 2004 (to be submitted).

Additional publications not included in this thesis are shown below.

Journal Papers

[221] M. B. Stegmann and R. Larsen. Multi-band modelling of appearance. *Image and Vision Computing*, 21(1):61–67, jan 2003.

[204] M. B. Stegmann. Analysis of 4D cardiac magnetic resonance images. *Journal of The Danish Optical Society, DOPS-NYT*, 4:38–39, dec 2001.

Conference Papers

[73] S. Darkner, R. Larsen, M. B. Stegmann, and B. K. Ersbøll. Wedgelet enhanced appearance models. In *2nd International Workshop on Generative Model-Based Vision – GMBV, CVPR 2004*, 2004 (to appear).

[141] R. Larsen, K. B. Hilger, K. Skoglund, S. Darkner, R. R. Paulsen, M. B. Stegmann, B. Lading, H. Thodberg, and H. Eiriksson. Some issues of biological shape modelling with applications. In J. Bigün and T. Gustavsson, editors, *13th Scandinavian Conference on Image Analysis (SCIA), Gothenburg, Sweden*, volume 2749 of *LNCS*, pages 509–519. Springer, jun 2003.

[111] D. W. Hansen, J. P. Hansen, M. Nielsen, A. S. Johansen, and M. B. Stegmann. Eye typing using Markov and active appearance models. In *IEEE Workshop on Applications of Computer Vision - WACV*, pages 132–136, dec 2002.

[112] D. W. Hansen, M. Nielsen, J. P. Hansen, A. S. Johansen, and M. B. Stegmann. Tracking eyes using shape and appearance. In *IAPR Workshop on Machine Vision Applications - MVA*, pages 201–204, dec 2002.

[117] K. B. Hilger, M. B. Stegmann, and R. Larsen. A noise robust statistical model for image representation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2002, 5th Int. Conference, Tokyo, Japan*, volume 2488 of *LNCS*, pages 444–451, sep 2002.

[220] M. B. Stegmann and R. Larsen. Multi-band modelling of appearance. In *First International Workshop on Generative Model-Based Vision – GMBV, ECCV 2002*, pages 101–106, Copenhagen, Denmark, jun 2002. DIKU.

[139] R. Larsen, H. Eiriksson, and M. B. Stegmann. Q-MAF shape decomposition. In Wiro J. Niessen and Max A. Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2001, 4th International Conference, Utrecht, The Netherlands*, volume 2208 of *LNCS*, pages 837–844. Springer, 2001.

[116] K. B. Hilger and M. B. Stegmann. MADCam – the multispectral active decomposition camera. In *Proc. 10th Danish Conference on Pattern Recognition and Image Analysis, Copenhagen, Denmark*, volume 1, pages 136–142. DIKU, 2001.

[205] M. B. Stegmann. Object tracking using active appearance models. In *Proc. 10th Danish Conference on Pattern Recognition and Image Analysis, Copenhagen, Denmark*, volume 1, pages 54–60. DIKU, 2001.

[214] M. B. Stegmann, R. Fisker, and B. K. Ersbøll. Extending and applying active appearance models for automated, high precision segmentation in different image modalities. In *Proc. 12th Scandinavian Conference on Image Analysis – SCIA 2001*, volume 1, pages 90–97, 2001.

[215] M. B. Stegmann, R. Fisker, B. K. Ersbøll, H. H. Thodberg, and L. Hyldstrup. Active appearance models: Theory and cases. In *Proc. 9th Danish Conference on Pattern Recognition and Image Analysis, Aalborg, Denmark*, volume 1, pages 49–57. AUC, 2000.

Technical Reports

[219] M. B. Stegmann and D. D. Gomez. A brief introduction to statistical shape analysis. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, mar 2002.

[207] M. B. Stegmann. An annotated dataset of 14 cardiac MR images. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, apr 2002.

[208] M. B. Stegmann. An annotated dataset of 14 meat images. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, apr 2002.

1.4 Mathematical Nomenclature

The mathematical notation used in this thesis is enumerated below to ease reading and understanding.

Vectors are formatted in columns and typeset in non-italic, lower-case, boldface using spaces to separate elements: $\mathbf{v} = [\, a\ b\ c\,]^{\mathsf{T}}$

Vector functions are typeset in non-italic boldface: $\mathbf{f}(\mathbf{v}) = \mathbf{v} + \mathbf{v}$

Matrices are typeset in non-italic, boldface, capitals: $\mathbf{M} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

Matrix diagonals are manipulated using the $\mathrm{diag}(a)$ operator. If a is a vector of length n, $\mathbf{a}$, an $n \times n$ diagonal matrix is produced. If a is an $n \times n$ matrix, $\mathbf{A}$, the diagonal is extracted into a vector of length n.

Dot-product operator (inner product) is typeset using a central dot: $\mathbf{a} \cdot \mathbf{b} = \sum_i a_i b_i$

Hadamard product (element-wise multiply) is denoted by the $\odot$ symbol.
E.g. for vectors: $\mathbf{a} \odot \mathbf{b} = \mathrm{diag}(\, \mathrm{diag}(\mathbf{a})\mathrm{diag}(\mathbf{b})\,)$

Sets are typeset using curly braces: $\{\alpha\ \beta\ \gamma\}$ or $\{\mathbf{x}_i\}_{i=1}^{N}$

Vectors of ones are typeset as: $\mathbf{1} = [\, 1\ \cdots\ 1\,]^{\mathsf{T}}$

Identity matrices are typeset as: $\mathbf{I} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$

CHAPTER 2

Medical Background

This chapter presents a brief introduction to certain medical issues. This should ease the understanding for readers without a background in medicine. Further, the chapter also gives a simplified introduction to the predominant image modality treated in the later chapters; magnetic resonance imaging (MRI).

2.1 Cardiac Nomenclature

This section enumerates the acronyms used for cardiac function and anatomy in the following chapters.

AMI	Acute myocardial infarction
CHF	Congestive heart failure
ED	End diastole
EDV	End diastolic volume
EF	Ejection fraction
endo	Endocardial contour
epi	Epicardial contour
ES	End systole
ESV	End systolic volume
HLA	Horizontal long axis
LA	Long axis
LV	Left ventricle
LVM	Left ventricular mass
RV	Right ventricle
SA	Short axis
VLA	Vertical long axis

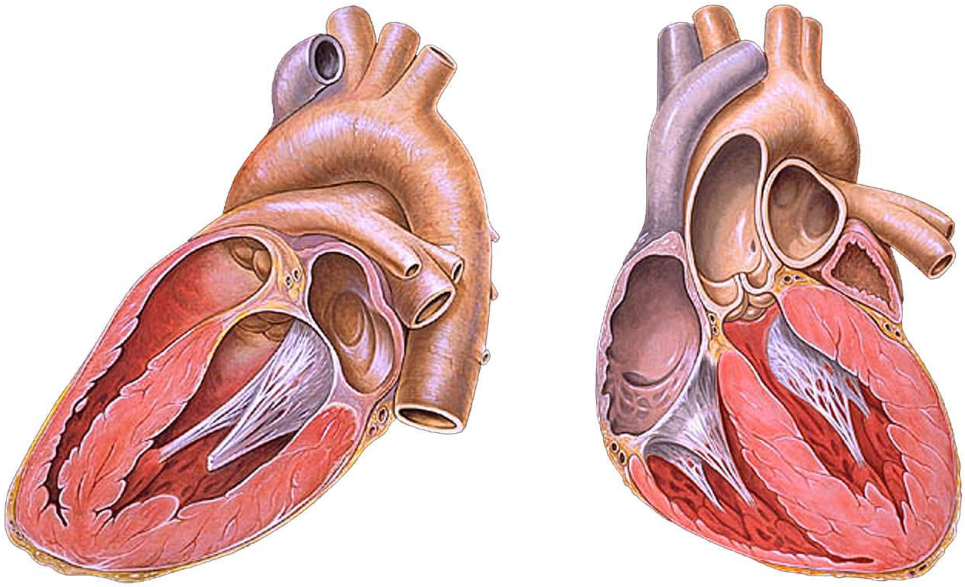


Figure 2.1: Cardiac anatomy. Two cross-sections along the left ventricular long axis. The left ventricle is shown to the right in both images and conversely; the right ventricle to the left. Illustration courtesy of Patrick J. Lynch, Yale University School of Medicine.

2.2 Cardiac Function and Anatomy

The human heart is essentially a pump that causes the transportation of deoxygenated blood to the lungs and return oxygen-rich blood to the cells in the body. It is divided into four chambers: the left and right atria, and the left and right ventricles. The right ventricle transports blood to the lungs while the – substantially stronger – left ventricle serves the whole body. This difference in muscle mass is clearly illustrated in Figure 2.1.

Although this thesis treats several issues, emphasis is put on analysis and quantification in conjunction with two major cardiac diseases: congestive heart failure (CHF) and acute myocardial infarction (AMI).

Congestive heart failure is a disorder in which the heart looses its ability to pump blood efficiently. The most common causes of CHF are hypertension (high blood pressure) and coronary artery disease. CHF is quantified by the *ejection fraction* (EF), which denotes the fraction of the end-diastolic blood volume expelled from the ventricle with each systole. In other words it measures the normalised stroke volume, which is the blood volume difference between the maximally relaxed heart (at end-diastole) and the maximally contracted heart (at end-systole), normalised with the end-diastolic volume, EDV:

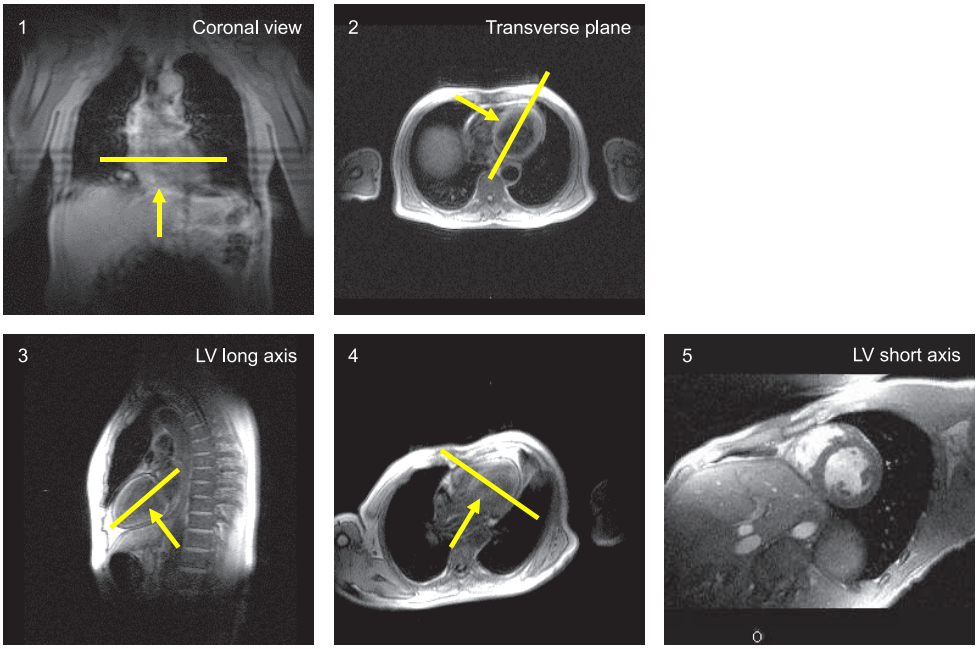


Figure 2.2: Short axis scan planning. Each image constitutes a step towards the standardised short axis scan plane of the left ventricle.

$$EF = \frac{EDV - ESV}{EDV}. \tag{2.1}$$

Here ESV denotes the end-systolic volume. EF constitutes the most commonly used parameter of systolic function in clinical pratice [86]. An EF above 55% is considered normal, while an EF below 40% is a clear indicator of CHF.

Chapter 13 deals with quantification of CHF by introducing a novel method for rapid and unsupervised estimation of EF. This is done from cardiac magnetic resonance imaging (MRI), which is considered the gold standard for cardiac imaging [170]. All cardiac chapters deal with short axis (SA) cardiac MRI, which is a special orientation of the MR image plane providing a standardised coordinate system intrinsic to the left ventricle. Due to voxel anisotropy, short axis images are optimal to quantify phenomena orthogonal to the long-axis only. This anisotropy is due to the limitations and scan time considerations inherent to MR acquisition of dynamic processes as the cardiac cycle. Figure 2.2 illustrates the process of manual short axis MR scan planning (a process, which was recently automated [72, 129, 145, 148]).

From a set of short axis images, a complete three-dimensional image of the heart can be compiled (an issue treated further in Chapter 12). This is illustrated in Figure 2.3, which is augmented with labels of anatomical areas referred to later on.

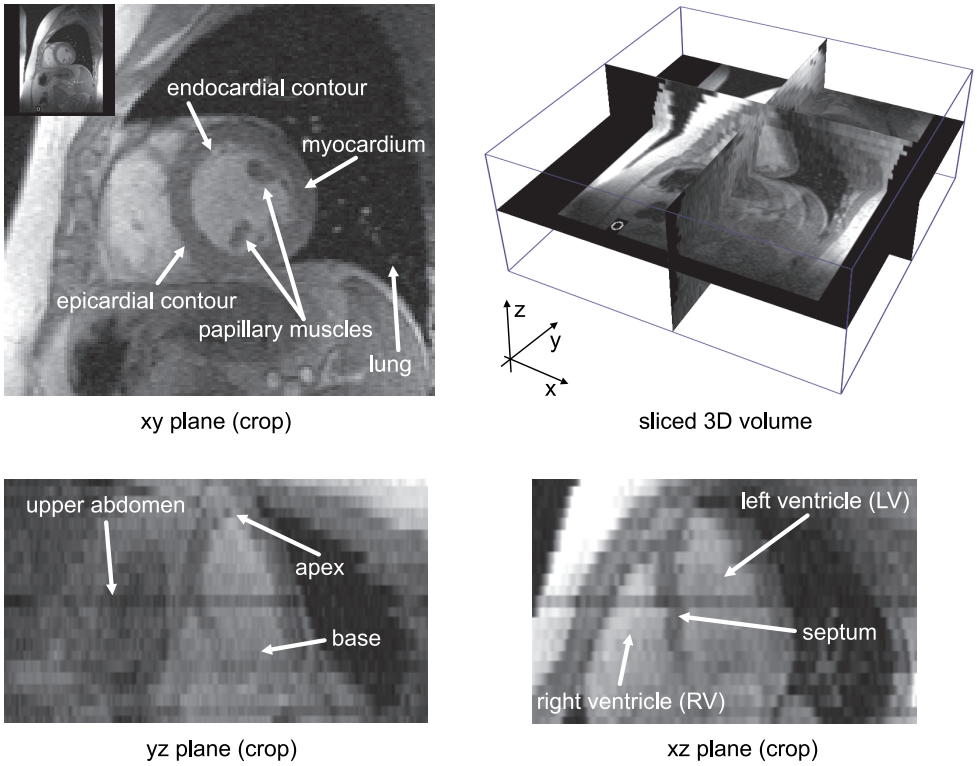


Figure 2.3: Cardiac anatomy depicted by MRI. Short axis image, SA (xy plane); sliced 3D volume showing the plane relationship; approx. vertical long axis image, VLA (yz plane); approx. horizontal long axis image, HLA (xz plane). Notice that the long-axis is inverted, i.e. a reflection in the apical-basal direction is seen in all images.

Notice that the anisotropic voxels results in an approximately four times lower through-plane resolution (yz/xz) compared to the in-plane resolution (xy). Further anatomical labels and definition of standardised cardiac views can be found in [46].

The second cardiac disorder dealt with in this thesis is acute myocardial infarction (AIM). This occurs when an area of the myocardium (the heart muscle) dies or is permanently damaged due to inadequate blood supply (perfusion). This is mostly caused by a clot blocking one of the coronary arteries, which serves blood to the myocardium.

To quantify cardiac perfusion MRI can be employed. By injecting a bolus of a paramagnetic contrast substance areas of the myocardium served by diseased arteries will show a delayed and attenuated response. Myocardial perfusion MRI may be carried out under the influence of physical or pharmacological stress, since ar-

eas affected by a coronary artery lesion may not exhibit a perfusion deficit under resting conditions [86]. Regional perfusion responses can be collected by volumetric myocardial perfusion MR sequences and analysed provided that each voxel will correspond to the same tissue part throughout the time-series.

Chapter 11 presents a method that establishes these aforementioned voxel correspondences by unsupervised motion-compensation of multi-slice perfusion MRI.

2.3 Magnetic Resonance Imaging

This section introduces the very basic principles of magnetic resonance imaging (MRI). The presentation is primarily based on internal material from the Danish Research Centre for Magnetic Resonance and the introductions given in [2, 122].

> *"You know, what these people do is really very clever. They put little spies into the molecules and send radio signals to them, and they have to radio back what they are seeing."*
>
> Attributed to physicist Niels Bohr about the principles behind magnetic-resonance imaging (MRI) [233].

Although in a popular form and pertaining to the none-imaging precursor to MRI, nuclear magnetic resonance (NMR), this quote describes the leading principle behind MRI quite well.

MRI relies on the magnetic properties of nuclei with an uneven atomic mass or atomic number. These nuclei possess a spin angular momentum, which induces a magnetic field coincident with the axis of the spin. Normally, the axes of such nuclei are entirely random. However, in the presence of an external magnetic field, nuclei will experience a torque, which causes them to precess around the axis of the external field. This is analogous to a spinning top in the gravitation field of the earth. The rate of precession, ω, has a very simple relation to the strength of the external magnetic field, $\mathbf{B}_0$, and the *gyromagnetic ratio*, γ, specific to the nucleus. This relation is expressed by the Larmor equation

$$\omega = \gamma|\mathbf{B}_0|. \tag{2.2}$$

For the commonly occurring hydrogen nucleus, ^{1}H, we have $\gamma \approx 42$ MHz/T.

The macroscopic alignment of nuclei causes a weak net magnetisation in direction of $\mathbf{B}_0$. There is no transverse component due to the stochastic phases of the nuclei. By emitting external radio frequency (RF) pulses orthogonal to $\mathbf{B}_0$, nuclei can now be brought out of their equilibrium state. Optimal disturbance is obviously reached at the resonance frequency. Thus, water (H_2O) in a 1.5 Tesla MR scanner is excited

Figure 2.4: MR scanner. Siemens Magnetom Avanto. Siemens press picture courtesy of Siemens AG, Munich/Berlin.

by an RF pulse of 63 MHz. The relaxation period where nuclei return to the equilibrium state will thus emit an RF pulse having the same frequency. Consequently, this pulse reveals the density of hydrogen nuclei related to the proportion of water molecules in the sample. Elaborated a little further, nuclei relaxation times are measured along two directions. The longitudinal (parallel to $\mathbf{B}_0$) time constant, T_1, which measures the recovery of the magnetic field due to the aligned nuclei, and the transversal time constant, T_2, which measures the decaying of the magnetic field orthogonal to $\mathbf{B}_0$ after excitation.

The above technique is called NMR and has been known long before MRI (see [233] for an interesting piece on this topic). Fortunately, it was discovered in the early 1970's – and recently Nobel Prize awarded – how magnetic gradient fields cunningly applied to the main magnetic field would allow for spatial encoding of the nuclei response by modifying their resonance frequencies. Decoding this frequency representation then carries out the image formation in MRI. Fortunately, the T_1 and T_2 mentioned above take on different values intrinsic to different tissue

types. This has lead to MRI acquisition protocols designed to enhance or suppress selected tissue types with extensive image contrast options yielding a hitherto unsurpassed versatility within medical imaging.

This concludes the introduction to magnetic resonance imaging; an essential and highly flexible method for non-invasive investigation of the human body in conjunction with cancer, cardiovascular diseases, multiple sclerosis, rheumatoid arthritis, et cetera.

An example of a modern MR scanner is given in Figure 2.4.

CHAPTER 3

Snakes and other Creatures

This chapter provides a brief introduction to early deformable models as well a sophisticated composite generative framework of today. The latter is the model class investigated in this thesis; Active Appearance Models.[1]

In recent years, the model-based approach towards image interpretation named *deformable template models* has proven very powerful. This is especially true in the case of noisy images and images containing objects with large variability. As definition of a deformable template model we will use the one due to Fisker [95]:

> **DEFINITION 3.1** *A deformable template model can be characterized as a model, which under an implicit or explicit optimisation criterion deforms a shape to match a known object in a given image.*

The earliest deformable template models date back to Widrow's 'rubber mask', and Fischler and Elschlager's 'spring-loaded' templates, both introduced in 1973 (see e.g. [92]). However, it should be fair to say that the most well known deformable template model is the Active Contour Model – also called *Snakes*, which was introduced by Kass et al. [134] in 1988. Snakes represent objects as a set of outline landmarks upon which a correlation structure is forced to constrain local shape changes. In order to improve specificity, many attempts at hand crafting prior knowledge into deformable template models have been carried out. These include the parameterisation of a human eye using ellipses and arcs by Yuille et al. [255].

In a more general approach, yet while preserving specificity Cootes and Taylor [57] proposed the Active Shape Models (ASMs), in which structural relationships and shape variability are learned through observation. The first ASM paper was published in 1992 and borrowed a fair amount of attention from the Snakes paper [134], by choosing the title *Active Shape Models – 'Smart Snakes'* [57]. This af-

[1] This chapter includes parts of Section 13.3.1.

filiation was muted in the later and major introduction *Active Shape Models – their training and application* [68] by T. F. Cootes, C. J. Taylor, D. H. Cooper and J. Graham, published in 1995.

The brotherhood with Snakes was rightfully claimed as they both are deformable models but contrary to Snakes, ASMs have global constraints on shape deformation. These constraints are learned through observation giving the model flexibility, robustness, and specificity as the model only can synthesize plausible instances similar to the observations. In practice, this is accomplished by a training set of annotated examples, which are aligned by a generalized Procrustes analysis [107, 108] followed by a principal component analysis (see Appendix 14.D).

A direct extension of the ASM approach has lead to the Active Appearance Models [52, 54, 88]. Besides shape information, image appearance is taken into consideration. This means that every pixel intensity across the object, is included into the model and subsequently used during the model-to-image match.

Jain et al. [130, 131] classifies deformable template models as either being *free form* or *parametric* where the former denotes model deformation dependent on *local* constraints on the shape and the latter *global* shape constraints. By building statistical models of shape and texture variation from a training set, AAM qualifies as being a parametric deformable template model.

Quite similar to AAMs and developed in parallel is the Active Blobs/Active Voodoo Dolls proposed by Isidoro and Sclaroff [128], Sclaroff and Isidoro [188]. Active Blobs is a real-time tracking technique, which captures shape and appearance information from a prototype image using a finite element method (FEM) to model shape variation. Compared to AAMs, Active Blobs deform a static texture, whereas AAMs optimises both texture and shape during image search.

Also based on a prototype – and a finite element framework using Galerkin interpolants – is the Modal Matching technique proposed by Sclaroff and Pentland [189]. Objects are matched using the strain energy of the FEM. A major advantage is that the objects can have an unequal number of landmarks and it easily copes with large rotations. Other concurrent work having resemblance to AAMs include the Multidimensional Morphable Models by Jones and Poggio [133] and the Eigen Tracking work by Black and Jepson [24].

For further information on deformable template models in general, the reader is referred to the reviews given in [25, 95, 130, 156].

3.1 A Brief Introduction to Active Appearance Models

Interpretation by synthesis has been shown to be a very powerful approach to image analysis. On a coarse level, this approach consists of two parts; i) a mathematical model, which is able to mimic the image formation process, and ii) a search regime, which is able to match this model to an image by letting the image synthesised by

the model – in some sense – be "similar" to the unseen image.

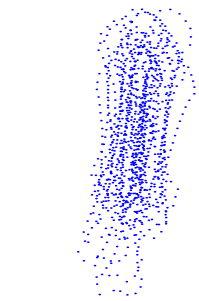
For such a strategy to be useful, certain requirements must be met for both parts. The model must either encode specific features of interests and/or contain parameters that reflect interesting latent variables of the objects in question. For example this could be point features locating interesting structures, and a variable related to the age of the object. Additionally, this model should only be capable of synthesising valid image instances of the object class in question. Matching such a model to an unknown image would thus draw inference about these properties. The matching procedure should provide a meaningful model-to-image match in a reasonable amount of time, dependent on the application, or alternatively reject the presence of the sought-after object.

Active Appearance Models (AAMs) [52, 54, 88] represent one method that has gained considerable attention in the literature, seeking to meet the above requirements. The encoded properties of the model are correlated movement of landmark points and latent texture variables controlling image appearance normalised for shape changes given by landmark points. In essence a fruitful marriage between the ideas of Eigenface models [237], Point Distribution models [57] and Active Shape Models [67] with some ingenious refinements.

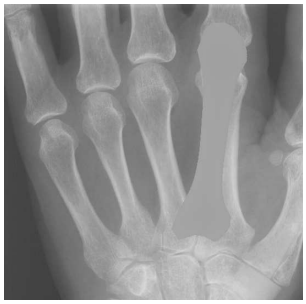
Formally, AAMs establish a compact parameterisation of shape and pixel intensities, as learned from a representative training set. The latter is also denoted *texture*. Objects are defined by marking up each example with points of correspondence (i.e. landmarks) over the training set either by hand, or by semi- to completely automated methods. Using a learning-based optimisation strategy, AAMs can be rapidly fitted to unseen images.

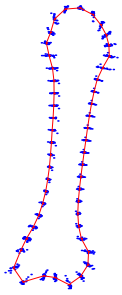
Figure 3.1 shows a pictorial of the AAM training process and AAM image search process in a case study of metacarpal bone two in hand radiographs. This case is treated in more depth in [203]. Figure 3.1.A shows the scatter of the landmarks of a set of 23 metacarpal outlines annotated using 50 landmarks (3.1.C). These shapes are subsequently aligned in 3.1.B. Synthetic shape examples from the resulting shape model are shown in Figure 3.1.F. Next, a model spanning the texture variation is built by sampling all training examples (similar to 3.1.E) using the Delaunay triangulation of the mean shape shown in 3.1.D. Synthetic textures in the mean shape configuration are shown in 3.1.G. To obtain a combined parameterisation of shape and texture variability, the two above models are coupled as shown in 3.1.H. These parameterised images of simultaneous shape and texture variation can then be employed to search new unseen images as shown in 3.1.1–3 where the model image is overlaid in its current configuration. In this case it took the iterative matching scheme 11 iterations to produce the final result shown in 3.1.3 and 3.1.4. The texture model contained 25487 intensity samples and the image search took 174 ms.

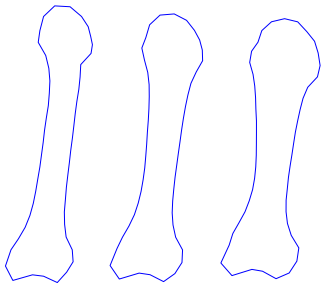
Notation and further AAM details are given in Chapter 7. Refer to [52, 54, 65] (and [203] as a supplement) for a level of description suitable for implementation.


A - Unaligned shapes.

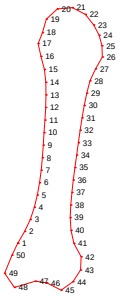

E - Annotated training image.

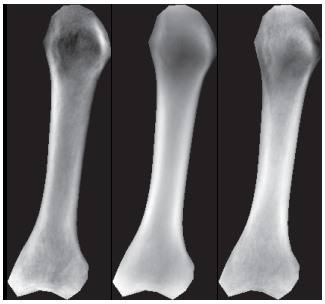

1 - Initial model configuration.


B - Aligned shapes.

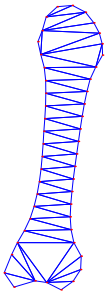

F - First mode of shape variation.

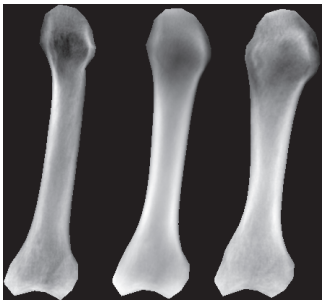

2 - After 5 iterations.


C - Landmarks on mean shape.


G - First mode of texture variation.


3 - Converged (11 iterations).


D - Delaunay triangulation.


H - First mode of combined variation.


4 - Converged shape.

Figure 3.1: Pictorial of the AAM training process (A–H) and AAM image search process (1–4) in metacarpal hand radiographs. All deformations are -3 std.dev., mean, +3 std.dev.

CHAPTER 4

Survey of Developments in Active Appearance Models

This chapter reviews the developments within generative modelling related to Active Appearance Models. Generative, and otherwise related, approaches only having a minor theoretical overlap with AAMs are omitted from this presentation. Instead the reader is referred to the recent reviews in [9, 84, 99, 118, 154].

First, an overview of the variations of – and extensions to – the original AAMs are presented grouped by subtopics.

This section is followed by a summary of medical AAM applications. Notice that the vast body of work on medical applications of the precursor to AAMs; Active Shape Models [57, 67, 68] is not included in the following.

Related approaches will only be mentioned in passing here. These include the Active Blobs by Sclaroff and Isidoro [188] and the Morphable Models by Jones and Poggio [133]. Refer to Cootes and Taylor [65] for further details.

The otherwise important body of work on determination of landmarks; either semi-automatically or fully automated is not considered an integral part of AAMs and is thus excluded from this chapter.

Both sections in this chapter comprise a thoroughly revised and extended edition of the summary given in Section 7.2.

4.1 Advances in Methodology

To ease of the overview of advances in AAMs it is useful to deconstruct the original formulation into a set of sub elements. This is carried out below along with the techniques that comprised each element in the initial papers by Edwards et al. [88] and Cootes et al. [52].

- Shape representation (landmark points)
- Shape alignment (2D generalised Procrustes alignment)
- Shape modelling (de-correlation by principal component analysis, PCA)
- Image warping (piecewise affine warp using a triangular mesh)
- Texture representation (normalised grey-scale intensities)
- Texture alignment (1D 'generalised Procrustes alignment')
- Texture modelling (de-correlation by principal component analysis)
- Combined modelling (principal component analysis)
- Search regime (iteratively, driven by principal component regression)
- Model truncation (based on variance)
- Model domain (2D, modelling one shape per image)

Alternative approaches to each element are given below along with suitable references to the relevant literature. Notice that some of the excellent scholarly work on AAM face modelling not pertinent to medical applications, and thus outside the scope of this chapter, has been left out.

Shape Representation

The conventional shape model used in an AAM is a point distribution model (PDM), which was introduced by Cootes et al. [67] and used in the later Active Shape Models [57, 68]. Interestingly, the precursor to PDMs was inter-point distance models via PCA due to Cootes et al. [49, 50]. Later, Heap and Hogg [113] demonstrated increased specificity and compactness in PCA modelling of Cartesian and polar coordinates in a synthetic data set.

We will leave the many variations here, noticing that any shape representation with fixed dimensionality (e.g. control points of a B-spline) is a suitable representation for PDM-inspired shape modelling.

Shape Alignment

Shape alignment in the L_2-norm has gained considerable attention in the literature (see e.g. Dryden and Mardia [80], Gower [108], ten Berge [231]). However, dependent on the data type, it may be beneficial to work in other norms. For example, Larsen et al. [139] applied AAMs with shape models aligned using the L_1, L_2 and L_∞-norms to metacarpal x-rays.

Shape Modelling

Even for complex biological phenomena, principal component analysis typically yields a very good decomposition of shape variability in a cohort. However, significant non-linearities exist in some cases, which render the implicit assumption of a multivariate Gaussian distribution invalid. Thus, PCA models will yield a poor specificity, leading to potential synthesis of implausible shape configurations. Some of these problematic cases are designed synthetically to emphasise the limitations of a PDM, while others are demonstrating actual, real-world examples of shape variability with dominating non-linearities.

Attempts to deal with such non-linearity include the polynomial regression PDM, PRPDM, by Sozou et al. [199]. Later, Sozou et al. [199] outperformed this using a back propagation neural network employing a multi-layer perceptron, which resulted in another xPDM acronym; the MLPPDM. A different approach is to employ a kernel-based density estimation of the shape distribution. This was proposed by Cootes and Taylor [60, 61] along with a computationally more attractive variant using a Gaussian mixture model to approximate the density function. Building on similar ideas Heap and Hogg [114] proposed a hierarchical PDM, the HPDM, also based on multiple Gaussian models. Non-linear shape models are also treated in depth by Bowden [38].

Advances within machine learning that allow working implicitly in infinite dimensional spaces, using so-called kernel methods, have also been utilised in shape modelling. Using this variant of non-linear PCA called Kernel PCA (KPCA) complex non-linear shape distributions can be modelled. This was demonstrated on shapes from projections of varying-angle faces by Romdhani et al. [186]. Further developments of this work was presented by Twining and Taylor [239] on synthetic shapes, and shapes from images of nematode worms.

PCA decomposes variation by maximisation of variance, which is easily shown using Lagrange multipliers. However, other measures may be of interest when a shape deformation basis is to be chosen. For example, Larsen [137, 138], Larsen et al. [139] chose to maximise the autocorrelation along 2D shape contours using the Maximum Autocorrelation Factors (MAF) due to Switzer [229]. This MAF approach was later extended to three-dimensional PDMs by Hilger et al. [115], Larsen and Hilger [140], Larsen et al. [142]. Interestingly, it turns out that Molgedey-Schusters algorithm for performing Independent Component Analysis (ICA) [165] is equivalent to MAF analysis, see [139].

Recently, ICA was reintroduced for shape modelling by Üzümcü et al. [241] with emphasis on the ordering of independent components. This was later incorporated into an AAM and evaluated on 2D cardiac MRI by Üzümcü et al. [242].

Image Warping

In the original AAM formulation dense image correspondences were established by image sampling via a shape-normalised (or *shape-free*) triangular mesh. Fixed sampling points on this mesh were propagated using Barycentric coordinates to a mesh similar in structure, but deformed in shape. At its best this constitutes a *homeomorphism*; a continuous and invertible deformation field. However, warp degeneracy arises if triangle normals are inverted, e.g. if two vertices of a triangle stay fixed and the third is mirrored. Fortunately, for high-quality meshes and moderate shape deformations this rarely happens. To this end, triangular meshes are typically established by a Delaunay triangulation of the Procrustes mean shape.

Nonetheless, certain applications do model structures that are more prone to these adverse effects. They include the chest radiographs of Chapter 10 and the capillary images treated by Rogers [185]. In the latter, three image warping strategies were evaluated; Delaunay mesh, manually defined mesh, and lastly, thin-plate splines [30]. Although all have their merits, none ensures a one-to-one mapping.

Recently, construction of invertible and C^∞ differentiable warp fields, so-called *diffeomorphisms* have appeared in the image warping literature, see e.g. [173, 238]. However, the current formulations do not fit well into performance-conscious frameworks such as AAMs, due to their excessive computational requirements.

Lastly, we mentioned that AAMs using piecewise affine image warping fit well into contemporary graphics hardware available on nearly any standard PC. This is explored in Chapter 7 and in more detail by Stegmann [206]. Ahlberg [3] also described a similar approach, developed independently.

Texture Alignment

In [52], image samples were compensated for scale and offset in an iterative approach similar to a generalised Procrustes analysis, obviously without rotation compensation. Texture vectors were linearised by a tangent space projection carried out by scaling, so that all texture vectors lie in the tangent hyper plane to the shape manifold, at the pole denoted by the Procrustes mean shape.

Later, Bosch et al. [33, 37] pre-processed texture vectors sampled from inherently noisy ultrasound images using histogram matching to let these approximate normally distributed samples more closely.

Texture Representation

AAMs were introduced as a generative model capable of synthesising near photo-realistic *grey-scale* images. This was soon extended by Edwards et al. [87] to texture vectors containing RGB tuples with applications to colour face modelling.

Later, Cootes and Taylor [64] demonstrated another multi-band texture representation aimed at reducing the sensitivity to lighting conditions. This had also applications to face modelling. Texture vectors comprised of two-tuples containing a local edge orientation estimate along with an edge reliability estimate.

Stegmann and Larsen [220, 221] also targeted sensitivity to lighting conditions in face modelling when exploring a composite representation of intensity, hue and edge strength.

Recently, Scott et al. [190] introduced another n-tuple texture representation consisting of reliability estimates of three low-level image features; edges, corners and gradients. Results were given on lateral spinal DXA scans and face images.

While all the above texture representations added to the size of the AAM texture vector, efforts in the opposite have also been carried out.

Primarily aimed at reducing computational costs Cootes et al. [51] used a subsampling scheme to reduce the texture model by a ratio of 1:4. The scheme selected a subset based on the ability of each pixel to predict corrections of the model parameters.

Later, Wolstenholme and Taylor [249, 250, 251] introduced the concept of wavelet-compressed AAMs by applying the Haar wavelet to AAMs of transversal brain MRI.

Stegmann and Forchhammer [217], Stegmann et al. [218] continued these studies where two wavelet bases were evaluated on face images along with a simple weighting scheme, which allowed for substantial compression with a simultaneously improvement in registration accuracy. Parts of this work are also described in Chapter 8.

Texture Modelling

As for shape modelling, PCA was also used to decompose texture variation by seeking a new basis spanning maximal variance. However, inspired by developments in shape modelling, Hilger et al. [117] explored a Maximum Autocorrelation Factors (MAF) representation of texture. This was carried out in a case study of cardiac MRI.

Alternative texture modelling was also applied by Stegmann and Larsson [223, 224], Stegmann et al. [227]. Here, an ensemble of linear texture models approximated the texture density function in order to cope with the highly non-Gaussian intensity distributions inherent to cardiac perfusion MRI time-series. Parts of this work are also described in Chapter 11.

Search Regime

Fitting an appearance model to an unseen image was initially based on iteratively applying additive model parameter corrections estimated from the result of a principal component regression. This was carried out by regressing observed texture differences between the model and the image, with known model parameter displacements [52, 88].

This approach was superseded by a more tractable approach using numeric differentiation as introduced by Cootes et al. [54], Cootes and Taylor [63].

In between these two developments, Cootes et al. [51] proposed the Shape AAM, in which texture parameters are made implicit by employing projections of image textures. Consequently, Shape AAMs only update shape and pose parameters during search.

AAMs based on principal component regression and numeric differentiation were treated and evaluated by Stegmann [206], Stegmann et al. [213]. Further, this comparative study comprises parts of Chapter 7.

Independently and simultaneously, Cootes and Kittipanya-ngam [56] made the same comparison, and additionally introduced a hybrid of the two. They also compared the four fundamentally different AAM search regimes; Basic AAM, Shape AAM, Compositional AAM (due to Baker and Matthews [12]), and the Direct AAM (due to Hou et al. [123]).

The compositional approach introduced in [12] is investigated in great detail in numerous publications; Baker et al. [10, 11], Baker and Matthews [13, 14], Matthews and Baker [155] holding many interesting views on image search using AAMs and related approaches. This is a very comprehensive body of work and a summary in this context cannot justify it. Refer directly to the references for further details.

Recently, Batur and Hayes [17] introduced an adaptive strategy for correction of model parameters updates based on the current texture configuration. The method was evaluated on face images with variable lighting conditions.

Stegmann and Davies [211], Stegmann et al. [212, 214] used the general purpose optimiser BFGS as a black-box method to refine on AAM search results. This provides a simple method to improve on, and validate, AAM convergence, albeit this may not be near as efficient as many of the above principled alternatives. This strategy has also been taken in Chapter 9 and Chapter 10.

Model truncation

Traditionally, PCA models in AAMs are truncated either by retaining a fixed number of principal components, or based on a variance constraint. The latter is typically carried out so that the number of principal modes retained is capable of spanning 95% of the variance present in the training set. This issue was treated by Stegmann

Table 4.1: AAM model domain (ordered chronologically).

Coupled 2D images	2D time series	3D images	Coupled 3D images
Cootes et al. [71]	Mitchell et al. [159]	Beichel et al. [20]*	Stegmann [210]
Lelieveldt et al. [147]	Lelieveldt et al. [146]	Mitchell et al. [162]	(Chapter 13)
Oost et al. [174, 175]	Sonka et al. [198]	Bosch et al. [35]*	
	Bosch et al. [33, 34, 34, 36]	Beichel et al. [19]*	
	Stegmann and Larsson [223]	Lelieveldt et al. [144]	
	Stegmann and Larsson [224]	Mitchell et al. [161]	
	Stegmann et al. [227]		
	(Chapter 11)		

*Either application, or model is not entirely three-dimensional.

et al. [213] (see Chapter 7) where parallel analysis [121] – a purely data-driven method for model truncation – was introduced in AAMs and evaluated.

Model domain

A natural extension of the idea of exploiting coherent information in 2D images, is to couple multiple views, 2D time-series or even increase the dimensionality of the model domain. This has been explored in numerous publications, which are categorised chronologically (top down) in Table 4.1.

A precursor to this development was the view-based AAMs introduced by Cootes et al. [70].

Additional Advances

Finally, a few topics outside the above taxonomy deserve mention. One area concentrates on relaxing the – occasionally too rigid – constraints imposed by the shape model, typically due to training sets being not representative enough. This important issue has been treated by Cootes and Taylor [58, 59, 62] and Roberts et al. [184].

Another advance, which at first glance might seem as just another application, is the disease characterization on basis of AAM parameters. This work was recently reported by Mitchell et al. [163] and Sonka et al. [197] and is essential an unleashing of the interpretative power inherent to AAMs.

Finally, the maximum-a-posteriori (MAP) formulation of AAMs by Cootes and Taylor [63] should be mentioned, which for example allows for including priors on point positions in conjunction with semi-guided search.

4.2 Medical Applications

The potential of AAMs in medical image analysis has become evident to many researchers since the beginning of this thesis work.

To give a concise presentation of the medical applications explored so far a grouping of scholarly work by modality and model is given in Table 4.2. References within each group are ordered chronologically.

Table 4.2: Medical Applications of AAMs.

Modality	Model	Reference
Brain MRI	Ventricles, caudate nucleus, lentiform nucleus	Cootes et al. [48, 53] Wolstenholme and Taylor [250] Cootes and Taylor [62, 65, 66]
Brain MRI	Corpus callosum	Stegmann and Davies [211] Stegmann et al. [212] (Chapter 9)
Knee MRI	Cartilage and bone	Cootes and Taylor [65]
Cardiac MRI	Left and/or right ventricle (SA, 3D, 3D+t)	Mitchell et al. [160, 164] Stegmann [203] Stegmann et al. [214, 226] Mitchell et al. [159] Lelieveldt et al. [146] Sonka et al. [198] Mitchell et al. [162] Lelieveldt et al. [144] Stegmann and Forchhammer [217] Hilger et al. [117] Mitchell et al. [161] Stegmann et al. [213] (Chapter 7) Stegmann et al. [222] Stegmann and Larsson [225] (Chapter 12) Stegmann [210] (Chapter 13)
Cardiac MRI	Left ventricle (SA+VLA)	Lelieveldt et al. [147]
Cardiac Perfusion MRI	Left and right ventricle	Stegmann and Larsson [223, 224] Stegmann et al. [227] (Chapter 11)
Ultrasound (echocardiograms)	Left ventricle	Bosch et al. [33] Bosch et al. [33, 34, 34, 36]
Hand radiographs	Metacarpals	Stegmann [203] Larsen et al. [139] Thodberg [234]
Chest radiographs	Lung, clavicles, heart	van Ginneken et al. [245] (Chapter 10)
Computed tomography (CT)	Diaphragm dome	Beichel et al. [19, 20]
Dual-energy x-ray absorptiometry (DXA)	Vertebrae	Scott et al. [190] Roberts et al. [184]
Electron microscope	Nerve capillaries	Rogers [185]
X-ray angiograms	Left ventricle (bi-plane)	Oost et al. [174, 175]

CHAPTER 5

Contributions

This thesis has contributed by giving solutions to a set of relevant problems in medical image analysis. All of these have been formulated based on the same theoretical framework, which is well described in the literature. Although focusing solely on sub problems is an integral part of technical research, this thesis has also emphasised on another aspect. Studies and developments of complete systems have been presented that – once trained – have been capable of solving their designated task autonomously. These tasks consisted of highly varying problems spanning different modalities and dimensionalities (2D, 2D+time, 3D/3D+time).

Figure 5.1 shows the different AAM applications and modalities treated. From top left: 2D cardiac MRI, 2D chest radiograph, cross-section of 3D cardiac MRI, deformation modes of 2D faces, 2D corpus callosum in brain MRI, frame from multi-slice cardiac perfusion MRI time series, model search in 4D cardiac cine MRI.

The fact that very different domain problems all can be cast into the same generic framework, has many benefits. Adaptations may be needed, as detailed in Part II, but relying on common ground means that all projects will benefit from general improvements and extensions. Secondly, and very important for the dissemination of the presented methods, as these build upon a well-known and well-described framework chances of a successful re-implementation and/or continuation by others are far greater. This should obviously always be a key success parameter in scholarly work.

Chapter 7 presented the core framework upon which the results of the following five chapters were based on. This software framework has been contributed to the medical image analysis community as an open, free and documented application programmers interface under the name *The AAM-API* alongside supporting example programs, scripts, reference data sets, reference measures, et cetera. Refer to Appendix 14.A for a brief, self-contained introduction to this API.

While most papers are centred around a particular medical application, Chapter 8 focuses on reducing computational and memory requirements of AAMs. Re-

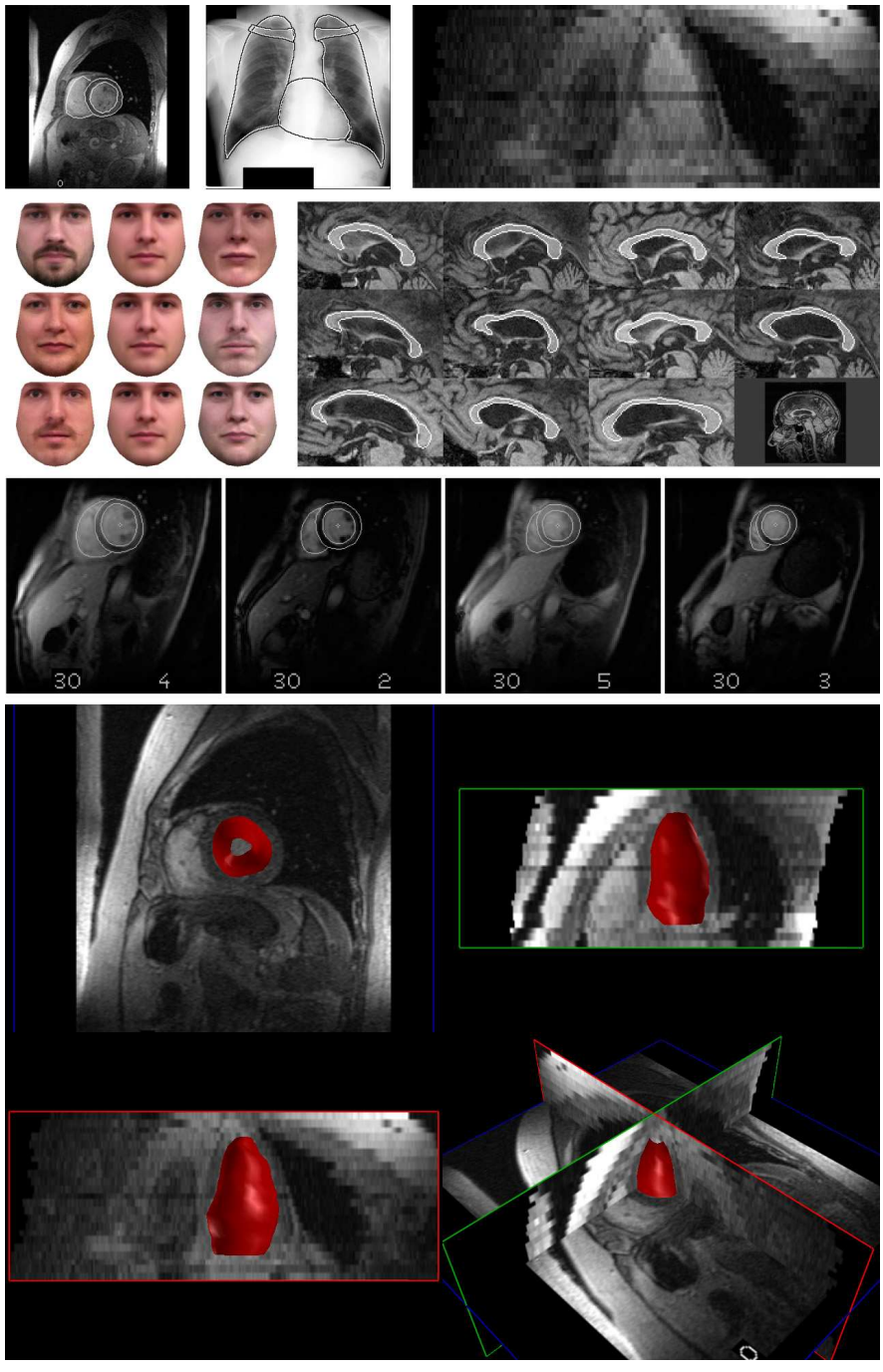


Figure 5.1: Collage of the different AAM applications and modalities treated in this thesis.

sults are demonstrated on 2D face images with an ambition to extend this to the very information-rich modalities often met in medical imaging. This work is based on the idea of applying wavelet compression to the AAM texture models by Wolstenholme and Taylor. Encouraging results were obtained when evaluating two different wavelet bases employing a weighting scheme exploiting the decomposition scheme used.

The corpus callosum study in Chapter 9, the chest radiographs in Chapter 10, and the perfusion MRI study in Chapter 11 all present problems that have been explored earlier in the medical image analysis literature. The major contribution wrt. application lies in demonstrating a generic framework capable of providing similar (or better) results with the very desirable property that they were all produced without any supervision.[1] AAMs have not previously been applied to these problems. Further, the combination of added benefits over a naïve segmentation, which include the interpretive powers of AAMs, the established dense correspondences and inherent shape-compensated representation has presumably not been offered previously to these applications.

A noteworthy aspect of Chapter 11 was the demonstration of the ability to handle pathologies. Several patients exhibited severe perfusion deficits, which were handled robustly by the presented framework. This is obviously very desirable but may not always be the case, an issue which will be touched upon later.

Fully automated methods for two additional applications were presented in Chapter 12 and Chapter 13. The former treated the problem of obtaining truly three-dimensional data from slice time-series of cardiac MRI acquired using breath-hold. This topic seems to have gained very little interest, if any, in the literature previously. The chapter contributes with a discussion of the problem followed by a solution designed for rapid and automated application.

Chapter 13 introduced a novel bi-temporal 3D AAM framework – utilising the results of Chapter 12 – where the triangles of 2D AAMs were substituted with tetrahedrons (among other issues). An efficient software implementation, independent from the previous API, was evaluated on 4D cardiac cine MRI. This spatio-temporal model may arguably represent the most sophisticated application of AAMs to date, judged from its dimensionality. It may also represent the fastest fully automated method for left ventricular ejection fraction estimation in 4D MRI. However, this has not been fully confirmed. Due to the limited size of the training set, the wavelet compression of Chapter 8 was not required in this study although it makes for an obvious extension in larger studies.

Revisiting the objectives of the thesis, it is emphasised that all of the above studies contributed with both qualitative and quantitative assessments of their efficacy.

Refer directly to the papers in Part II for detailed contributions with respect to methodology. A condensed presentation of these is given in Section 1.2.

[1] Except for the single initialisation failure mentioned in Chapter 11.

CHAPTER 6

Summary

This chapter presents a discussion of the thesis work in a broader perspective and touches some of the relevant future challenges. Concluding remarks are given at the end of this chapter.

6.1 Discussion

The work presented in this thesis has revolved around the Active Appearance Models. In summary, good results have been obtained in many very different applications. Unfortunately, this is not a universal fact. Not all cases fit well into this scheme. First of all, AAMs need consistent features. For amorphous 'blobs', or cases with changes in topology the approach taken here is not suitable. Further, if the image quality is sufficiently high, less committed approaches such as level sets and fast marching [192] are much more appropriate. As mentioned earlier, cases exhibiting large shape variations may also tax the image warping methods used in AAMs beyond reason. However, this may be considered a technicality in the implementation, rather than an inherent flaw of AAMs.

Another obvious criticism of training-based methods such as AAMs is the tedious task of training set generation. Although very relevant, this issue has more sides to it. Establishment of ground truth solutions to the problem at hand is a necessary evil. Clinical validation is an invariable prerequisite prior to actual usage of any method within medical image analysis. The real problem is that methods such as AAMs may have more strict requirements than the typical manual solution of a problem. For example, if only a cross-sectional area of a brain structure is required, clinicians will not bother with landmarks, although such are the basis of AAMs. Fortunately, recent landmarking methods address this discrepancy by establishing points of correspondence from naïve segmentations. This was demonstrated in Chapter 9 using an MDL-based solution by Davies et al. [78], which is

also extended to three dimensions elsewhere [75, 77]. An alternative to this is the B-spline-based landmarking [98, 101], which is exemplified on 3D left and right ventricular shape models in MRI. Both of these approaches may form a basis for continuation of the work presented in Chapter 13, where the landmarking method used was rather crude. Using sophisticated landmarking techniques will not only alleviate the training set generation, but also allow for modelling of complex shape deformations. To this end, the used clinical convention of including the papillary muscles in the blood pool is far from optimal as this introduces a consistent bias in the results.

Modelling more detail inevitably requires more training data. This is the cost of having global priors. A vocal criticism of global models akin to AAMs is that they often are not able to generalise to new examples, e.g. with pathologies, different from the training set. Models having local priors similar to Snakes [134] or a level set formulation do not have this problem. They will easily deform to highly complex shapes found in the unseen image. Level sets can even deal with changes in topology. So, why use global priors? Because they ensure that only correct solutions can be reached by the model. (Provided that the model assumptions are met as treated in Chapter 11.) This is the key to the robustness demonstrated in this thesis. Solutions are to a great extent imputed to areas exhibiting weak, or hardly any, image evidence based on the overall shape and appearance. Examples include the noise study in Chapter 7, the fornix in the corpus callosum study in Chapter 9, the rather fuzzy blood/myocardium interface in some apical-basal slices of the cardiac MRI treated in Chapter 13, the anterior-posterior heart interface in the chest radiographs in Chapter 10, and the pre-bolus endocardial contours in Chapter 11. However, there is no such thing as a free lunch. Global priors require training data and models can trade generalisation ability for robustness. This is called the bias/variance trade-off and is inherent to all mathematical modelling [105]. This should be realised by researchers claiming that e.g. PDMs are flawed since they do not generalise to pathologies not seen in the training set. This is a feature, not a flaw. If the image evidence is unambiguous, this feature is not needed, contrary to the cases treated in this thesis.

At this point it is relevant to stress that due to the rapid development in medical imaging, the need for models with strong priors versus weak priors constantly should be reassessed on a per-application basis. Problems today that require the constrained bases provided by AAMs may well suffice with a more uncommitted solution tomorrow, e.g. due to a significant increase of image contrast/signal-to-noise ratio. This does not mean that models with strong priors in the long run will be superfluous. Imaging technology will remain to be pushed further in the pursuit of new – and previously unreachable – applications.

Hybrid methods capable of being highly constrained, but also provide gradual relaxation towards extreme flexibility seem to constitute a very fruitful area of research. A standard method for estimating this – preferably continuous – regularisation parameter is cross-validation. This will provide a sense of flexibility adaptive

to the given data quality in generic frameworks such as AAMs. However, it is not clear how such hybrid schemes should deal with extreme flexibility such as changes in topology.

Although being highly constrained, AAMs can appear as pure guesswork compared to other models. Biomechanical models, which extend to modelling of muscle fibre directions and the electrophysiology activation patterns demonstrates a completely different richness in detail, see e.g. the work on cardiac models by [176, 191, 196]. Having proper statistical models at this level would provide the ultimate tool for inference. Currently, this does not seem to be feasible, partly due to computational reasons and the many modalities and techniques involved in creating such models. Nevertheless, this development points towards a very interesting future, albeit on a long-term basis.

Methods based on non-rigid registration (NRR) of images share several properties with AAMs. Typically, NRR is based on choosing a representative image as an atlas where features of interest are marked up. These features are subsequently located in unseen images by a pair-wise matching of the atlas image and the unseen image. One popular realisation of such a framework is the combination of the uncommitted free-form deformation (FFD) B-spline basis and the normalised mutual information (NMI) [228] voxel similarity measure introduced in [187]. Although, no texture variation is modelled in this framework it is able cope with substantial differences due to the highly invariant normalised mutual information measure. Despite a hierarchical implementation and interpolation kernels with limited support, this scheme is computationally more demanding than AAMs. However, the tedious AAM training is avoided altogether in return (except for the atlas image). A first step towards applying this scheme to registration of 4D cardiac MRI is described in [178]. An extension of the scheme was applied to tagged and untagged 4D cardiac MRI in [182].

Research in combining these two paradigms by using NRR to build statistical models is ongoing, see e.g. [69]. This is aimed at combining the best of both worlds: the speed and inherent interpretive powers of AAMs with the significantly less demanding training process of NRR. However, issues regarding bias/variance and clinical validation still remain pertinent to such a unified framework.

6.2 Conclusion

This thesis has demonstrated frameworks which solves several important problems in medical image analysis. Substantial parts of the computational frameworks developed during the thesis period have been made publicly available[1], facilitating reproducibility and further developments. Emphasis has been put on providing quantitative analyses in addition to qualitative assessments.

[1] See the appendix (Chapter 14).

The presented methods and extensions have been fully automated and performed rapidly in the vast majority of the treated domain problems. The quality of the results has in some cases – despite limited training data – compared well to the typical inter-observer variability.

Only future clinical validation can document the true potential of the presented methods, but from this standpoint, and with respect to the treated applications, they all hold great promise for the days to come.

Part II

Contributions

CHAPTER 7

FAME – A Flexible Appearance Modelling Environment

Mikkel B. Stegmann, Bjarne K. Ersbøll, and Rasmus Larsen

Abstract

Combined modelling of pixel intensities and shape has proven to be a very robust and widely applicable approach to interpret images. As such the Active Appearance Model (AAM) framework has been applied to a wide variety of problems within medical image analysis. This paper summarises AAM applications within medicine and describes a public domain implementation, namely the Flexible Appearance Modelling Environment (FAME). We give guidelines for the use of this research platform, and show that the optimisation techniques used renders it applicable to interactive medical applications. To increase performance and make models generalise better, we apply parallel analysis to obtain automatic and objective model truncation. Further, two different AAM training methods are compared along with a reference case study carried out on cross-sectional short-axis cardiac magnetic resonance images and face images. Source code and annotated data sets needed to reproduce the results are put in the public domain for further investigation.

Keywords: active appearance models, face segmentation, left ventricular segmentation, public domain training data and software

7.1 Introduction

Modelling biological variation present in medical images has been a challenge for a long time. The variation is often very complex and each image modality has its own set of characteristic artefacts. While the amount of biological complexity is either fixed or very slowly changing, new imaging modalities emerge and existing are improved, often very rapidly. This accentuates the need for general methods capable of adapting to both issues above. In addition, we would prefer these to be reasonably fast, specific to the current problem and finally, robust to noise and acquisition artefacts.

During the past few years, the generative modelling framework: Active Appearance Models (AAM) [52, 88] has aimed at meeting all of the above requirements. By being capable of synthesising near photo-realistic images of objects, AAMs are taking the analysis-through-synthesis approach to the extreme. This approach has proven its worth in numerous different applications, both medical and non-medical.

We sincerely believe that progress is most easily obtained through *transparency*. By this we mean that algorithms and reference data set should be made available along with corresponding reference performance measures. In that way new and existing methods can be compared on a fair and transparent basis by eliminating, or at least significantly alleviate, the tedious process of re-implementing competing methods and re-generating similar training data, et cetera.

This paper represents a small step towards transparent research by presenting a public domain AAM implementation along with reference measures carried out on likewise public domain data. Focus is put on the different design choices an AAM designer must make prior to model building. In particular, we evaluate two different variations on the AAM formulation and treat the bias/variance problem occurring when selecting the number of principal components.

The paper is organised as follows: Section 7.2 summarises medical applications of AAMs. Section 7.3 serves as an introduction to the inner workings of AAMs. Section 7.4 describes the two prevalent methods for calculating model parameter updates. Section 7.5 treats the problem of selecting the number model parameters. Section 7.6 describes our AAM framework, The Flexible Appearance Modelling Environment (FAME). Section 7.7 describes two sets of training data and gives segmentation results using FAME. Finally, Section 7.8 serves a discussion and conclusion.

7.2 Background

At the introduction of Active Appearance Models [52, 88] focus was put on segmentation and interpretation of face images. However, being a generic approach to build models from an annotated training set, medical applications were soon to follow. This section summarises the use of AAMs in medical imaging, but the list is by

no means exhaustive. For example, the vast amount of literature on AAM-related variations and precursors has been omitted.

Both [48, 53, 62, 250] showed AAMs modelling ventricles, the caudate nucleus and the lentiform nucleus in single-slice magnetic resonance images (MRI) of the brain. This case was also treated in [66] and [65] where the latter also demonstrated an AAM of single-slice MRI of cartilage and bone in the knee. Up to this point medical AAM work was done by the inventors, primarily Cootes, Taylor and Edwards of Manchester University.

Mitchell et al. [160, 164] demonstrated the applicability of AAMs on single-slice cardiac MRI and later on time series of cardiac MRI [146, 159, 198]. Bosch et al. [33, 36] used a similar approach on time series of echocardiograms. A direct continuation of this work is the 3D AAMs built on volumetric cardiac MRI by Mitchell et al. [161, 162] and Lelieveldt et al. [144]. Two variations of 3D AAMs were applied to diaphragm dome CT by Beichel et al. [19, 20] and to echocardiogram time series by Bosch et al. [35]. Duchesne et al. [82] have used a variation of the traditional AAM, with a shape model similar to that of the Morphable Models of Jones and Poggio [133], for 3D segmentation of the hippocampus in brain MRI. Stegmann et al. [203, 214, 217, 226] used AAMs for segmentation of metacarpal radiographs and cardiac MRI. Later Thodberg [234] used a Shape AAM for segmentation of metacarpal radiographs. A modified AAM, using multiple texture models, was used for segmenting myocardial perfusion MRI sequences by Stegmann and Larsson [223, 224]. Hilger et al. [117] explored a noise robust texture representation on ventricular cardiac MRI. Finally, Stegmann et al. [211, 212] used sequential AAMs based on MDL-defined landmarks for corpus callosum analysis in brain MRI.

7.3 Active Appearance Models

Active Appearance Models establish a compact parameterisation of object variability, as learned from a training set by estimating a set of latent variables. The modelled object properties are usually shape and pixel intensities. The latter is henceforward denoted *texture*. From these quantities new images similar to the training set can be synthesised.

Objects are defined by marking up each example with points of correspondence (i.e. landmarks) over the set either by hand, or by semi- to completely automated methods. From these landmarks a shape model [68] is built. Further, given a suitable warp function a dense (i.e. per-pixel) correspondence is established between training objects, thereby enabling a proper modelling of texture variability. By exploiting prior knowledge of the nature of the optimisation space, these models of shape and texture can be rapidly fitted to unseen images, thus providing image interpretation through synthesis.

Variability is modelled by means of a Principal Component Analysis (PCA), i.e.

an eigenanalysis of the dispersions of shape and texture. Let there be given P training examples for an object class, and let each example be represented by a set of N landmark points and M texture samples. The shape examples are aligned to a common mean using a Generalised Procrustes Analysis (GPA) [108] where all effects of translation, rotation and scaling are removed. The obtained Procrustes shape coordinates are subsequently projected into the tangent plane to the shape manifold, at the pole given by the mean shape. The texture examples are warped into correspondence using a piece-wise affine warp and subsequently sampled from this *shape-free* reference. Typically, this geometrical reference frame is the Procrustes mean shape. Let $\mathbf{s}$ and $\mathbf{t}$ denote a synthesized shape and texture and let $\bar{\mathbf{s}}$ and $\bar{\mathbf{t}}$ denote the corresponding sample means. New instances are now generated by adjusting the PC scores, $\mathbf{b}_s$ and $\mathbf{b}_t$ in

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{\Phi}_s \mathbf{b}_s \tag{7.1}$$

and

$$\mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_t \mathbf{b}_t \tag{7.2}$$

where $\mathbf{\Phi}_s$ and $\mathbf{\Phi}_t$ are matrices of column eigenvectors of the shape and texture dispersions estimated from the training set. To obtain a combined shape and texture parameterisation, $\mathbf{c}$, the values of $\mathbf{b}_s$ and $\mathbf{b}_t$ over the training set are combined into

$$\mathbf{b} = \begin{bmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_t \end{bmatrix} = \begin{bmatrix} \mathbf{W}_s \mathbf{\Phi}_s^{\mathrm{T}} (\mathbf{s} - \bar{\mathbf{s}}) \\ \mathbf{\Phi}_t^{\mathrm{T}} (\mathbf{t} - \bar{\mathbf{t}}) \end{bmatrix}. \tag{7.3}$$

A suitable weighting between pixel distances and pixel intensities is carried out through the diagonal matrix $\mathbf{W}_s$ (see [65]). To make the normalised measures of pixel distance and pixel intensities commensurate, the shape PC scores are typically weighted by the square root of the ratio between the sums of the texture and shape eigenvalues.

To recover any correlation between shape and texture the two eigenspaces are usually coupled through a third PC transform

$$\mathbf{b} = \mathbf{\Phi}_c \mathbf{c} = \begin{bmatrix} \mathbf{\Phi}_{c,s} \\ \mathbf{\Phi}_{c,t} \end{bmatrix} \mathbf{c} \tag{7.4}$$

obtaining the combined appearance model parameters, $\mathbf{c}$, that generates new object instances by

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{\Phi}_s \mathbf{W}_s^{-1} \mathbf{\Phi}_{c,s} \mathbf{c} \quad , \quad \mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_t \mathbf{\Phi}_{c,t} \mathbf{c}. \tag{7.5}$$

To regularise the model and improve speed and compactness, $\mathbf{\Phi}_s$, $\mathbf{\Phi}_t$ and $\mathbf{\Phi}_c$ are truncated, usually such that a certain amount of variance in the training set is preserved. This eventually results in k ($k < P$) combined modes, i.e. k dynamic parameters encoded in the vector $\mathbf{c}$.

The object instance, $(\mathbf{s}, \mathbf{t})$, is synthesised into an image by warping the pixel intensities of $\mathbf{t}$ into the geometry of the shape $\mathbf{s}$. Given a suitable similarity measure the model is matched to an unseen image using an iterative updating scheme based on

a fixed Jacobian estimate [54, 65] or a principal component regression [52]. Typically the L_2-norm is used as similarity measure, which is sensible if the errors between the model and the image are (approximately) normally distributed.

This sums up the basic theory of AAMs. For further details refer to [52, 54, 65] (and [203] as a supplement).

7.4 AAM Training

Traditionally, AAMs have been trained to update model and pose parameters using one of two schemes described in the following. These parameter updates are carried out using difference images between the current model image and the corresponding part of the unseen image that it covers. Applying such parameter corrections in an iterative scheme should drive the model towards the ground truth shape in the image.

Multivariate Regression

The original AAM formulation [52, 88] uses a regression approach where difference vectors, $\delta\mathbf{t} = \mathbf{t}_{image} - \mathbf{t}_{model}$, are regressed onto corresponding parameter perturbation/displacement vectors, $\delta\mathbf{p}$. Here $\mathbf{p}$ is model and/or pose parameters, having the length Q. The goal is thus to obtain an optimal – in a least-squares sense – prediction matrix, $\mathbf{R}$, satisfying the linear relationship:

$$\delta\mathbf{p} = \mathbf{R}\delta\mathbf{t}. \tag{7.6}$$

Our formulation uses separate prediction matrices for model and pose parameters, denoted $\mathbf{R}_c$ and $\mathbf{R}_\rho$, respectively. Let there be conducted S perturbation experiments and let

$$\mathbf{P} = \begin{bmatrix} \vdots & & \vdots \\ \delta\mathbf{p}_1 & \dots & \delta\mathbf{p}_S \\ \vdots & & \vdots \end{bmatrix} \quad \text{and} \quad \mathbf{T} = \begin{bmatrix} \vdots & & \vdots \\ \delta\mathbf{t}_1 & \dots & \delta\mathbf{t}_S \\ \vdots & & \vdots \end{bmatrix}. \tag{7.7}$$

During these perturbation experiments $\delta\mathbf{t}_i$ is short for $\delta\mathbf{t}_i(\mathbf{p}^+ + \delta\mathbf{p}_i)$, where $\mathbf{p}^+$ denotes the optimal parameter configuration. Since $Q \ll S \ll M$ typically, $\mathbf{R}$ is estimated using principal component regression [65]. From the large matrix $\mathbf{T}$, an s-dimensional subspace ($s \leq S$) is extracted. This makes $\mathbf{R}$ well-determined in $\mathbf{P} = \mathbf{R}\mathbf{T}'$, where $\mathbf{T}'$ contains s-dimensional projected versions of $\mathbf{T}$. Consequently, an eigen-decomposition of an $S \times S$ matrix is involved. To keep the dimensionality of this matrix down to a feasible size, typically only a subset of the training set is used for training. This is especially important when estimating $\mathbf{R}_c$, as the number of model modes grows with the number of examples, P. For models with a small

number of training examples the growth in S becomes close to quadratic. In the experiments below the same training subset has been used to estimate both $\mathbf{R}_c$ and $\mathbf{R}_t$.

In the subsequent sections, this learning approach is denoted *Regression*.

The Fixed Jacobian Matrix Estimate

In later AAM publications [54, 63, 66] the multivariate regression is superseded by a simpler approach. It is easier to implement, faster to calculate and requires far less memory to execute. First we introduce the residual vector $\mathbf{r}$, parameterised by $\mathbf{p}$:

$$\mathbf{r}(\mathbf{p}) = \delta\mathbf{t}(\mathbf{p}) = \mathbf{t}_{image}(\mathbf{p}) - \mathbf{t}_{model}(\mathbf{p}). \tag{7.8}$$

A first order Taylor expansion of $\mathbf{r}$ at $\mathbf{p}^*$ is

$$\mathbf{r}(\mathbf{p}^* + \delta\mathbf{p}) \approx \mathbf{r}(\mathbf{p}^*) + \frac{\partial\mathbf{r}(\mathbf{p}^*)}{\partial\mathbf{p}}\delta\mathbf{p} \tag{7.9}$$

where $\mathbf{p}^*$ is in the proximity of $\mathbf{p}^+$ and

$$\frac{\partial\mathbf{r}(\mathbf{p}^*)}{\partial\mathbf{p}} = \frac{\partial\mathbf{r}}{\partial\mathbf{p}} = \begin{bmatrix} \frac{\partial r_1}{\partial p_1} & & \frac{\partial r_1}{\partial p_Q} \\ \vdots & \cdots & \vdots \\ \frac{\partial r_M}{\partial p_1} & & \frac{\partial r_M}{\partial p_Q} \end{bmatrix}. \tag{7.10}$$

The goal of the parameter update is to drive the residual vector to zero, i.e. finding $\mathbf{p}^+$. Using the L_2-norm the optimal $\delta\mathbf{p}$ is: $\arg\min_{\delta\mathbf{p}} ||\mathbf{r}(\mathbf{p}^* + \delta\mathbf{p})||^2$. Hence, the least-squares solution of (7.9) becomes:

$$\delta\mathbf{p} = -\left(\frac{\partial\mathbf{r}}{\partial\mathbf{p}}^{\mathsf{T}}\frac{\partial\mathbf{r}}{\partial\mathbf{p}}\right)^{-1}\frac{\partial\mathbf{r}}{\partial\mathbf{p}}^{\mathsf{T}}\mathbf{r}(\mathbf{p}^*) = -\mathbf{R}\mathbf{r}(\mathbf{p}^*). \tag{7.11}$$

To obtain optimal numerical stability, a singular value decomposition (SVD) of the Jacobian matrix, $\frac{\partial\mathbf{r}}{\partial\mathbf{p}}$ is preferred in order to obtain its pseudo-inverse, $\mathbf{R}$. However due to the size this is not feasible, which is why a normal matrix inversion must be carried out.

Normally, the Jacobian matrix must be recalculated at each $\mathbf{p}^*$, which is a computationally very expensive task due to its size. However, since AAMs operate in a standardised domain, i.e. the shape-free reference frame, AAMs perform the following approximation

$$\frac{\partial\mathbf{r}(\mathbf{p}^*)}{\partial\mathbf{p}} \approx \frac{\partial\mathbf{r}(\mathbf{p}^+)}{\partial\mathbf{p}}. \tag{7.12}$$

Further – and this is a somewhat crude assumption – the right-hand side of (7.12) is considered constant for all training examples. Thus, $\mathbf{R}$ is considered fixed and

estimated once during the model building process using numeric differentiation on the training set [54]. Let δp_{jk} denote the k-th perturbation of the j-th parameter, $\mathbf{e}_j$, a unit-vector for the j-th dimension and $w(\cdot)$, a weighting kernel, e.g. Gaussian [54]. Using a central estimator on all P training examples, the j-th column of the Jacobian is estimated as

$$\widehat{\frac{\partial \mathbf{r}}{\partial p_j}} = \frac{1}{P} \sum_{i}^{P} \sum_{k} w(\delta p_{jk}) \frac{\mathbf{r}(\mathbf{p}_i^+ + \delta p_{jk}\mathbf{e}_j) - \mathbf{r}(\mathbf{p}_i^+ - \delta p_{jk}\mathbf{e}_j)}{2\delta p_{jk}}. \tag{7.13}$$

In our experiments, the Jacobian matrix is estimated without weighting, i.e. using a uniform kernel, $w(\cdot)$.

In the subsequent sections, this learning approach is denoted the *Jacobian*.

Perturbation Scheme

The remaining design choice in both learning methods is the perturbation scheme. Although being a rather important step in crafting a good AAM, this topic is often not described in AAM literature.

For all experiments shown in this paper we used the perturbation scheme shown in Table 7.1. This has earlier been used over a wide range of cases with good results, see e.g. [117, 212, 223, 224]. However, the perturbation scheme remains a free parameter and could be subject for a thorough study on a per case basis to obtain optimal segmentation accuracy. To compare the two AAM training methods, given one perturbation scheme, each parameter is displaced separately, while the remaining parameters are kept at zero. Alternatively, as mentioned in [56], one could seek the optimal perturbation scheme given a specific training method. See the recent work in [56] for a variation of the AAM training, which allows simultaneous perturbation of all model parameters using both the regression and Jacobian training method.

Table 7.1: Perturbation scheme used in both learning methods.

Variable	Perturbations
x, y	$\pm 5\%$, $\pm 10\%$ of the width and height, respectively
θ	± 5, ± 15 degrees
$scale$	$\pm 5\%$, $\pm 15\%$
$c_1 \dots c_k$	± 0.25, ± 0.50 standard deviations

7.5 Model Truncation

Typically, the eigenspaces of shape, texture and combined variation are truncated so that each explains a fixed amount of variance. Since the variance along the i^{th} principal axis is equal to the corresponding eigenvalue, λ_i, this is easily carried out. To retain p percent of the variation, t modes can be chosen satisfying:

$$\sum_{i=1}^{t} \lambda_i \geq \frac{p}{100} \sum_{i} \lambda_i \tag{7.14}$$

How to estimate p is a classic bias/variance problem. Choosing a high value for p could result in a model too flexible, in the sense that it will be able to fit to noise present in the training set (high variance) and thus not generalise very well. Conversely, low p values would produce very restrictive models, not matching the training set very well (high bias towards the model). A common heuristic is to consider the remaining 5% of the signal to be noise.

An alternative method for choosing p is cross-validation, i.e. partitioning of the data set into a training and an independent test set. By adding modes the reconstruction error will initially decrease. But at some p the model will start to over-fit the training set and the reconstruction error on the test set will start to increase. To make this procedure less negatively biased and more robust several partitions can be used going towards a leave-one-out analysis in the extreme. More elaborate methods such as the bootstrap [89] could also be applied.

A convenient alternative to cross-validation is parallel analysis introduced by Horn [121], where the data is compared to either i) independent normally distributed synthetic samples, or ii) a randomised version of the data matrix.[1] We will concentrate on the latter since it is imposing less strict distributional assumptions on the data. Further, it is simpler to implement and calculate.

In short, parallel analysis seeks the amount of inflation in the eigenvalues due to sampling errors (also known as *chance capitalisation* or *chance correlation*). In the perturbation version of parallel analysis this is estimated by breaking the correlation between variables, i.e. for each variable the order of the observations is randomised. In the case of a shape model, this would be a randomisation of each landmark over all shapes. From a scree plot of the eigenvalues of the original data and the scrambled data, modes that have higher eigenvalues for the randomised data than the original data, can thus be considered noise. The rationale is that these modes are only stemming from chance capitalisation and not actual population correlation.

This data permutation version of Horn's parallel analysis is typically embedded in a Monte Carlo simulation scheme, where mean eigenvalues of several scrambling experiments are used to reduce sampling errors. Experiments are done with replacement due to the massive effort involved in keeping track of the permutations.

[1] The original work of John L. Horn used the former method.

7.6 FAME

The Flexible Appearance Modelling Environment (FAME) is a software framework containing an open source implementation of the Active Appearance Models. This core part of FAME is called the AAM-API. In addition, the FAME package provide *AAMLab* a GUI tool for image annotation, model exploration and model fitting, i/o scripts for communication with Matlab, test data sets etc. The complete package including data material, source code, documentation, scripts etc. needed to reproduce the results below are placed in the public domain and can be downloaded from:

`http://www.imm.dtu.dk/~aam/`

FAME is implemented in C++ and runs under Microsoft Windows. Two third-party libraries, VisSDK [158] and LAPACK [4] are required for image handling and eigen-decompositions. Both of these are also placed in the public domain.

Emphasis has been put on high performance and on providing a cornucopia of documentation features. This paper will present several examples of the latter. Due to the unsupervised analyses in AAMs, we believe that a partly exploratory approach to AAM building is required to craft good models.

Details of the implementation are placed in the appendix to provide a quick overview. Appendix 7.A provides observations on and changes from the original AAM formulation. Appendix 7.B enumerates the different design choices prior to model building using FAME.

AAMs rely heavily on image warps. It is often believed that image sampling and warping constitute the major part of an optimisation iteration. We have found that this is not the case. In the current implementation, warps can be carried out in optimised software code or by exploiting available OpenGL compliant graphics hardware. The latter is described in Appendix 7.C. To assess performance and the general applicability of the framework details on machinery and timings are given in Appendix 7.D.

Where the results below are not obtained directly from FAME, simple Matlab scripts were used. All of these had no parameters. Hence, this should not limit the reader from reproducing the results.

7.6.1 Extending FAME

FAME includes a command line interface called *aamc* that demonstrates basic usage. All experiments in this article were carried out using this. However, FAME users are encouraged to modify, extend or embed the framework in any way supporting education or research. We use FAME in two ways, i) from the ”outside” by calling methods etc., and ii) from the ”inside” by deriving new versions of the concepts using C++ inheritance.

7.7 Case Studies

This section comprises two cross-sectional case studies, where the first is an example of very complex biological variation in both shape and appearance. Moreover, it has the advantage that all readers have a visual system fine-tuned towards this case study. It consists of a cohort of face images. Examples of medical use of statistical face models include the work of Hutton et al. [125], which is aiming at a better understanding of syndromic facial dysmorphologies, here studying the Noonan syndrome.

The second case study addresses an invariable prerequisite for quantitative functional analysis of the human heart [99], namely segmentation of cardiac chambers. In this study the left and right ventricles are segmented in a set of single-slice cardiac MRI.

7.7.1 Face Images

The data set consists of 37 still images of 37 different frontal human faces, all without glasses and with a neutral expression. The gender distribution is 7 females and 30 males. Images were acquired in 640×480 JPEG colour format using a DV video camera. The following facial structures were manually annotated using 58 landmarks in total: eyebrows, eyes, nose, mouth and jaw. A total of seven point paths were used; three closed and four open. Refer to Figure 7.1 for an example annotation.

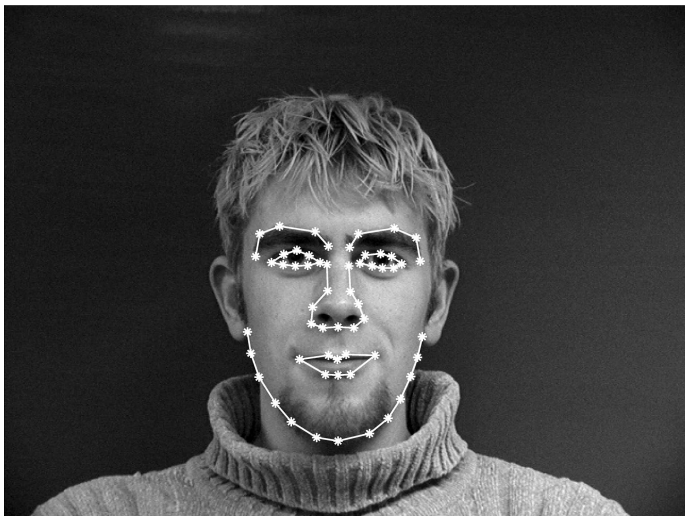


Figure 7.1: Example annotation of a face using 58 landmarks.

Shape Model

The foundation of the shape model is the 58 facial landmarks shown in Figure 7.2 (left). To establish a reference coordinate system relative to these landmarks within their convex hull a Delaunay triangulation is calculated and shown in Figure 7.2 (right). This constitutes the coordinate system of the texture model.

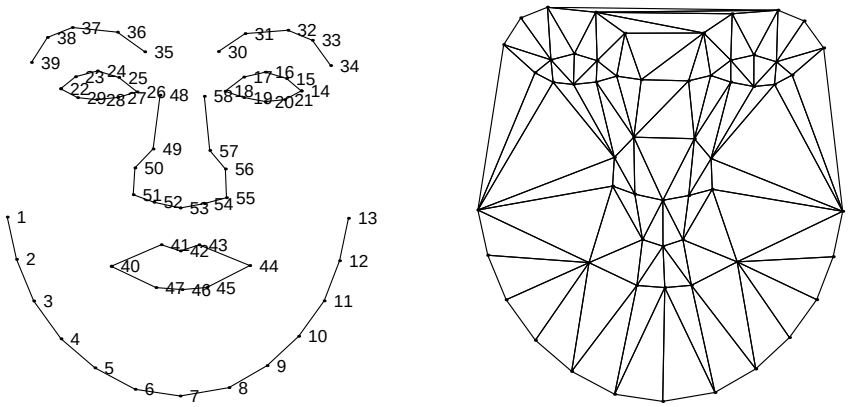


Figure 7.2: Facial landmarks (left). Triangular model mesh (right).

Plotting the scatter of all 37 face annotations yields the rather confusing plot in Figure 7.3 (left). To obtain only shape variation all Euclidean transformations have been filtered out in Figure 7.3 (right) by means of a Generalised Procrustes Analysis (GPA). Further, the Procrustes mean shape is fully drawn.

A better impression of landmark variation is given in Figure 7.4 (left), where the principal directions of each point are plotted. This reveals that the point variation of the aligned set of faces is often heteroscedastic, especially for landmarks at the upper jaw and on the lips.

To obtain an impression of landmark correlation, refer to Figure 7.4 (right) showing the canonical correlation[2] between each landmark. The five major block diagonals are jaw, eyes, eyebrows, mouth and nose.

Showing the three major modes of shape variation, Figure 7.5 displays the three directions with the highest variance in the subspace spanned by the 37 face annotations, which is embedded in a 2×58 dimensional space. The most dominant deformation is the upward movement of the upper jaw together with a downward movement of nose and mouth, and vice versa. When visualising this deformation as a movie sequence, it is clear that a major part of the variation picked up in this direction in the hyperspace does not correspond to inter-subject variation in head

[2]The maximal correlation between two sets of variables when general linear mappings on both are allowed.

Figure 7.3: Unaligned shapes (left). Aligned shapes (right).

anatomy. It is merely due to changes in the projection of the 3D landmarks into the 2D image plane, stemming from the head posture.

The amount of variance explained by the ten largest eigenvalues is shown in Table 7.2. While the three modes of Figure 7.5 covered 58% variation, it requires 20 modes to retain 95% of the total shape variation present in the training set.

To examine if any outliers are included into the shape model, all 37 faces are projected onto the first and second shape mode in Figure 7.6. In this face study, observation number 28 is revealed as an outlier in principal component two. However in this case, this underlines the limited training set size rather than indicating an abnormal jaw growth.

Finally, the principal scores of the five largest shape modes are inspected for non-

Table 7.2: Ten largest eigenvalues of the shape model.

Mode	Variance	Acc. variance
1	37.34%	37.34%
2	12.66%	50.00%
3	8.22%	58.22%
4	5.92%	64.14%
5	4.64%	68.77%
6	4.32%	73.10%
7	3.45%	76.55%
8	2.69%	79.24%
9	2.43%	81.67%
10	2.18%	83.85%

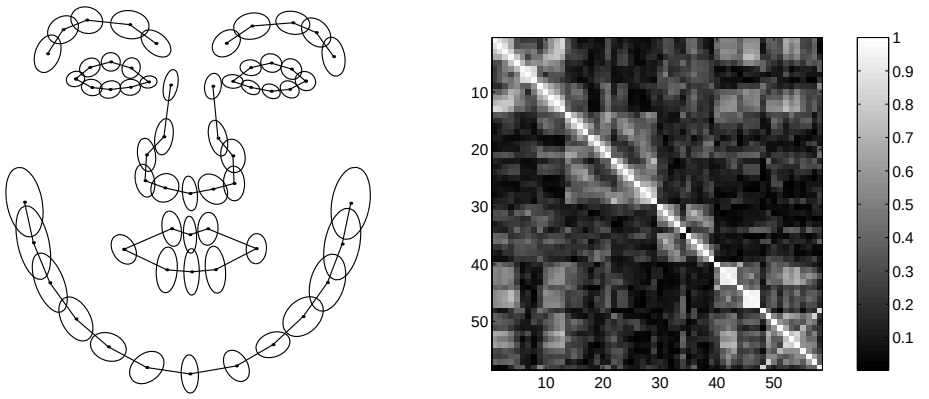


Figure 7.4: Principal point directions (left). The canonical correlation of landmarks (right).

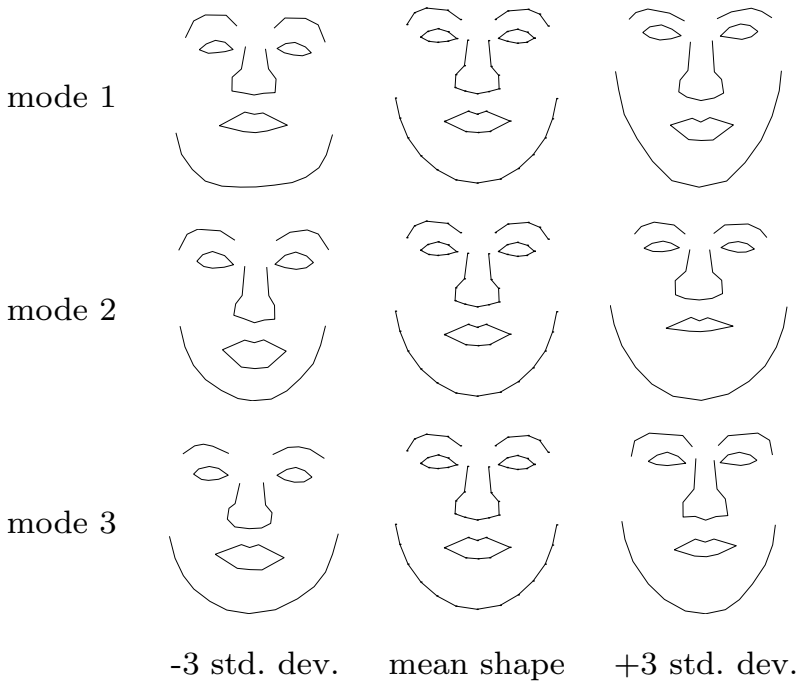


Figure 7.5: Shape deformation using the three largest principal modes (row-wise, top-down).

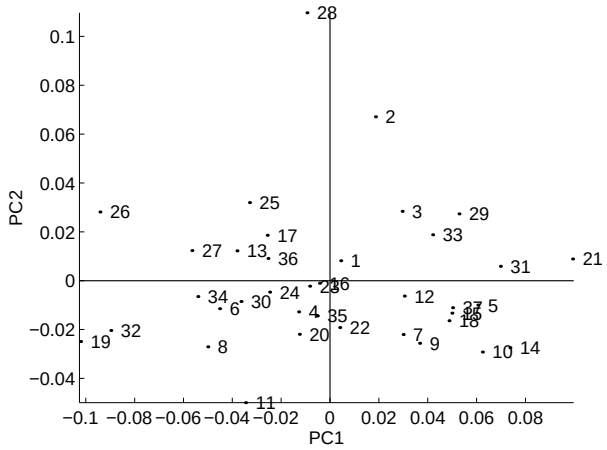


Figure 7.6: PC1 ($b_{s,1}$) versus PC2 ($b_{s,2}$) in the shape PCA.

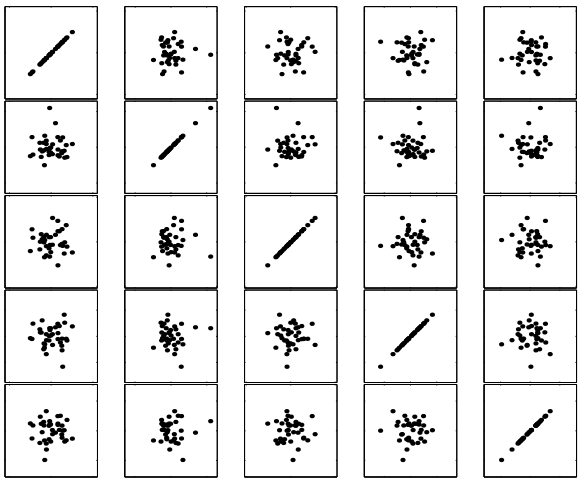


Figure 7.7: Scatter plot of the five largest shape principal components (PC1 is top-left). Axes are set to $\pm 4.5\sigma$ for each parameter.

linearities in the scatter plot in Figure 7.7 to ensure a proper (i.e. specific) shape representation. Figure axes are scaled accordingly to the variance of each parameter, thus depicting Gaussian distributions as circular symmetric densities. From this figure, the basis is accepted as not being over-complete, i.e. a generative multivariate Gaussian model will cover the face space sufficiently well and not generate implausible face shapes.

Texture Model

The face texture model for all 37 faces is built from 31221 sample positions in the reference frame, which is the Procrustes mean shape sized to mean size. The sample positions are obtained by sampling the reference frame in a grid aligned with the x and y axis and with one-pixel spacing between grid points. At each position a red, green and blue sample was obtained, resulting in a colour texture vector of 93663 samples.

Table 7.3 shows the ten largest eigenvalues of the texture PCA model. The three largest of these, accounting for 37% variation, are visualised as deformations of the mean texture in Figure 7.9, $\pm$ three standard deviations. Retaining 95% of the texture variation in the training set requires 29 modes. To inspect the training set for texture outliers all 37 faces are projected onto the first and second principal axes in Figure 7.8. From the first mode in Figure 7.9 we see that Figure 7.8 can be used to determine the degree of "beardedness" in faces in this case.

Table 7.3: Ten largest eigenvalues of the texture model.

Mode	Variance	Acc. variance
1	19.80%	19.80%
2	9.58%	29.38%
3	7.61%	36.99%
4	6.52%	43.51%
5	5.69%	49.20%
6	4.71%	53.91%
7	4.15%	58.06%
8	3.50%	61.56%
9	3.29%	64.84%
10	2.96%	67.81%

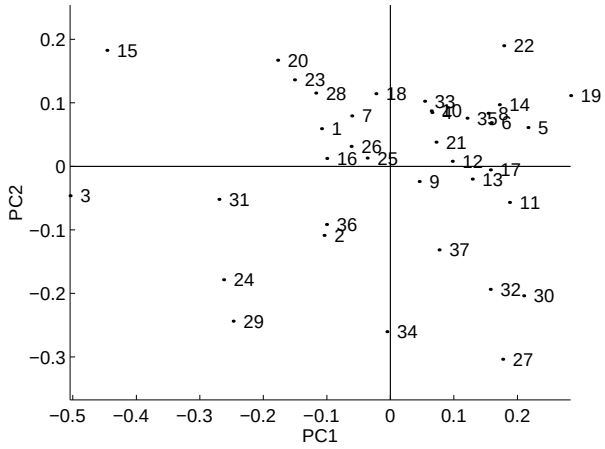


Figure 7.8: PC1 ($b_{t,1}$) versus PC2 ($b_{t,2}$) in the texture PCA (right).

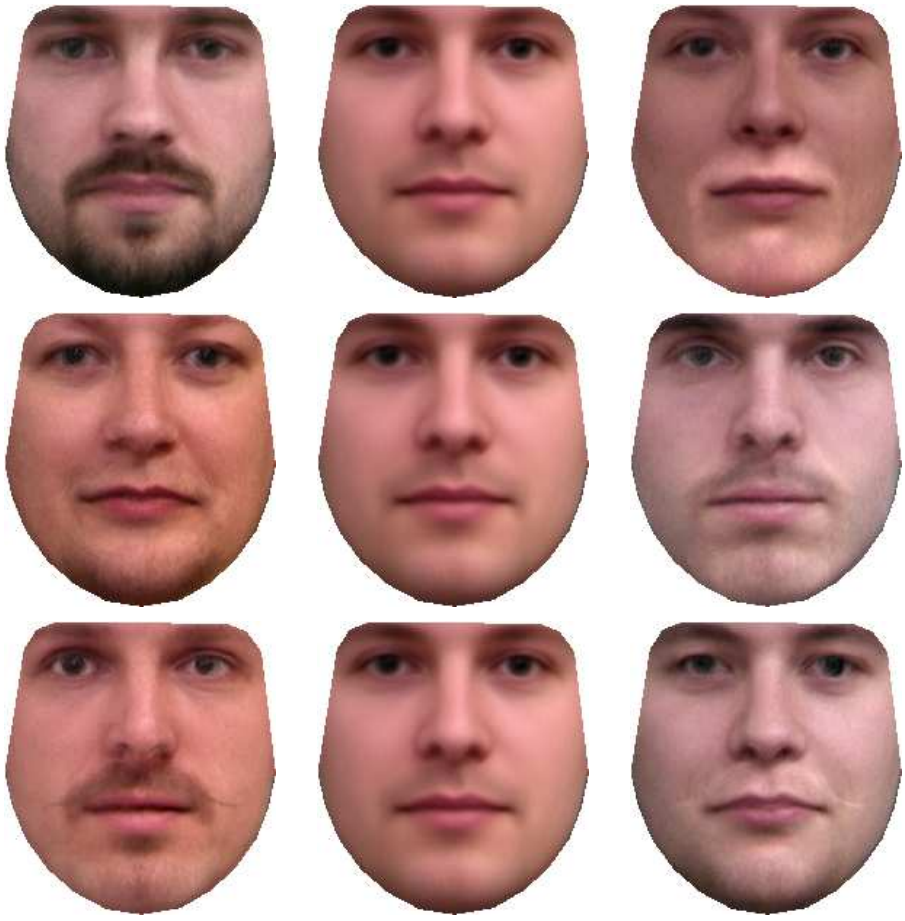


Figure 7.9: Texture deformation using the three largest principal modes (row-wise, top-down).

Appearance Model

Traditional AAMs use coupled eigenspaces of shape and texture by combining the PC scores in a third PC analysis. The ten largest eigenvalues are shown in Table 7.4. The corresponding three largest deformation modes of shape and texture are shown in Figure 7.10. Refer to FAME tool *AAMLab* for real-time exploration of the modes of the combined appearance model.

Table 7.4: Ten largest eigenvalues of the combined model.

Mode	Variance	Acc. variance
1	22.74%	22.74%
2	12.59%	35.33%
3	7.82%	43.16%
4	5.81%	48.96%
5	5.17%	54.13%
6	4.29%	58.42%
7	4.00%	62.42%
8	3.42%	65.84%
9	3.14%	68.98%
10	2.94%	71.91%

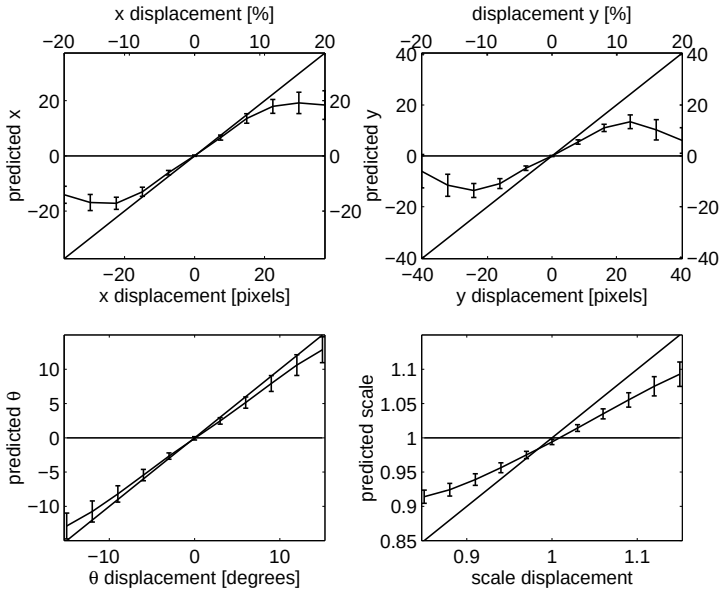
Model Training

In Section 7.4, two methods for model training were summarised. This section aims at assessing the quality of each of these methods for predicting AAM parameters. Both hypothesised a simple linear relationship between a residual vector and needed parameter updates. Desirable properties of such a prediction matrix include i) ability to predict the displacements learned, ii) ability to inter- and extrapolate the displacements learned, and most importantly iii) high prediction accuracy.

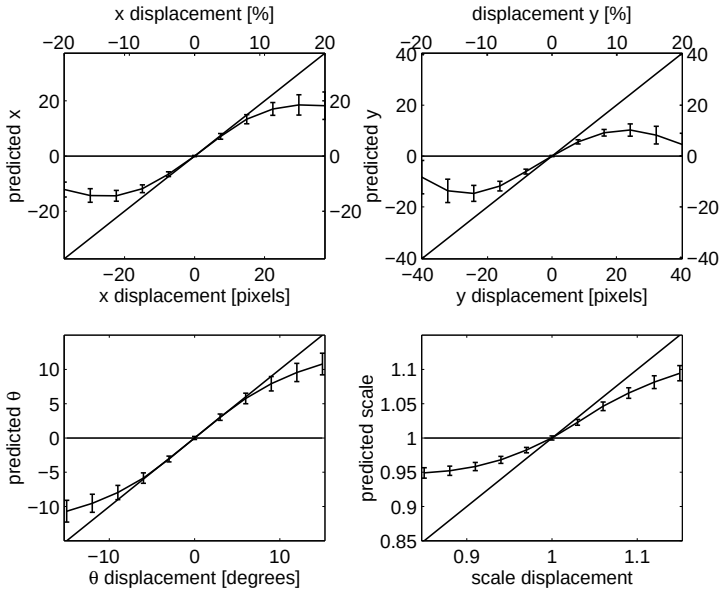
To assess the pose prediction abilities, all 37 training shapes were systematically displaced in all pose parameters, one by one. Results for both learning methods are shown in Figure 7.11. Error bars are one std. dev. In all cases only four images was used to train the models (the first image and then every tenth). These plots provide no significant evidence for choosing one learning method over the other. Consequently, we suggest the Jacobian due to its substantial lower computational complexity.

Due to the subsampling scheme, Figure 7.11 is a mixture of predictions upon known, $\frac{4}{37}$, and unknown data $\frac{33}{37}$. The next section will investigate how well these training methods generalise to unseen images with a mixture of displaced parameters.

Figure 7.10: Combined deformation using the three largest principal modes (row-wise, top-down).



(a) Colour, regression



(b) Colour, Jacobian

Figure 7.11: Prediction of pose parameter updates for large models.

Segmentation Accuracy

Though the generative properties of AAMs enable very sophisticated image interpretation directly from the model parameters, the most common application remains to be segmentation.

To assess the segmentation accuracy the model was initialised using the mean configuration displaced $\pm 10\%$ in x and y, relative to its width and height, i.e. four experiments per training example.

Due to the limited size of the training set, cross-validation was applied in a leave-one-out analysis leading to 37 models built from 36 examples for each evaluation. Double-mean landmark errors were calculated as the mean of all landmark points and mean of all leave-one-out experiments. Pt.pt. measures the Euclidean distance between corresponding landmarks of the model and the ground truth. Pt.crv. measures the shortest distance to the curve in a neighbourhood of the corresponding landmark. Thus, $4 \times 58 \times 37 = 8584$ pt.pt. and pt.crv. distances were calculated in each evaluation.

Results for colour and grey-scale face AAMs in two resolutions are shown in Table 7.5 and 7.6. AAMs in Table 7.6 were built in 1:2 decimated versions of the training images, i.e. 320×240 pixels. These models had on average 7799 and 23398 texture samples (grey-scale and colour, respectively). All shape, texture and combined PCA models were truncated to explain 95% of the variation present in the training set, resulting in 24 combined model parameters on average. From both tables we observe that the two training schemes do not differ significantly in performance. Actually, the Jacobian is slightly better, which is of utmost interest due to the far smaller computational and memory demands of the Jacobian training scheme.

From Tables 7.5 and 7.6 we also observe that the addition of colour to the models only resulted in a modest improvement of the final segmentation accuracy. However, most likely due to the gained specificity from adding colour, we have experienced that it also adds significant stability to the parameter update process (see [206]). Note the relatively small penalty in segmentation accuracy when working on 1:2 decimated images.

Table 7.5: Segmentation results – large models.

Model type	Learning method	Mean pt.pt.	Mean pt.crv.
Grey-scale	Regression	6.31±1.31	3.10±0.87
Grey-scale	Jacobian	6.13±1.39	2.98±0.95
Colour	Regression	6.13±1.08	3.09±0.84
Colour	Jacobian	5.93±1.15	2.86±0.86

Due to the excessive memory consumption of the regression approach, the subsampling scheme in Section 7.4 was rather crude. The memory usage of the Jaco-

Table 7.6: Segmentation results – small models.

Model type	Learning method	Mean pt.pt.	Mean pt.crv.
Grey-scale	Regression	3.31±0.83	1.64±0.52
Grey-scale	Jacobian	3.25±0.75	1.56±0.50
Colour	Regression	3.33±0.67	1.65±0.46
Colour	Jacobian	3.15±0.61	1.50±0.45

bian estimation does not depend on the number of training shapes. Consequently, we have tested if the more robust Jacobian estimate provided by using all training examples could result in higher segmentation accuracy. The results shown in Tables 7.7 and 7.8 support this by showing a subtle increase in accuracy.

Table 7.7: Segmentation results – large models, no subsampling.

Model type	Learning method	Mean pt.pt.	Mean pt.crv.
Grey-scale	Jacobian	6.03±1.37	2.91±0.96
Colour	Jacobian	5.92±1.11	2.88±0.83

Table 7.8: Segmentation results – small models, no subsampling.

Model type	Learning method	Mean pt.pt.	Mean pt.crv.
Grey-scale	Jacobian	3.14±0.69	1.51±0.46
Colour	Jacobian	3.08±0.59	1.48±0.41

Model Truncation

Using parallel analysis (PA) as presented in Section 7.5 the shape and texture models can be truncated according to the scree plots shown in Figure 7.12. In this experiment based on all 37 examples, the first 3 shape modes and the first 10 texture modes are selected. In contrast a 95% variance threshold would select 20 and 29 modes, respectively. As expected, we observe in the very dense and highly correlated texture models that the eigenvalues stemming from parallel analysis are very different from the data eigenvalues. And vice versa for the sparser and less correlated shape models. This is not a characteristic behaviour of texture and shape models. But due to AAMs ability to recover dense correspondences with a sparse set of landmarks, and the (often occurring) manual labour involved in placing landmarks, shape models are typically much more sparse than texture models.

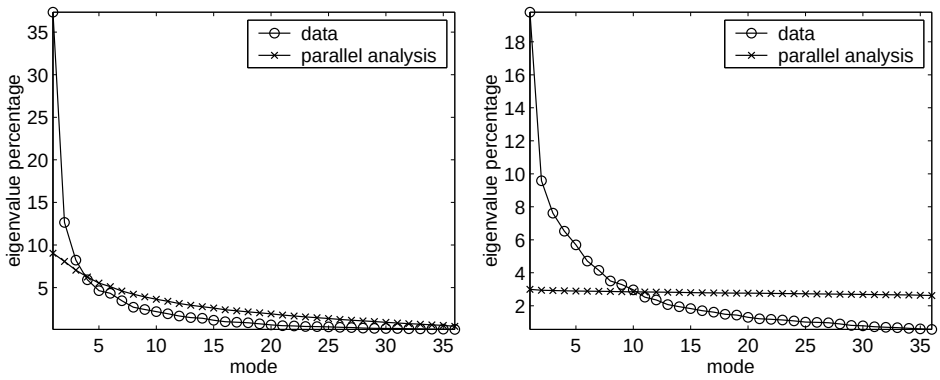


Figure 7.12: Scree plots for raw (o) and randomised data (x). Shape (left), texture (right).

Having a computational perspective, the simpler model is always preferable. However w.r.t. segmentation accuracy, it could prove to be too crude, i.e. resulting in under-fitting. On the other hand the variance threshold could be too liberal, resulting in over-fitting. We have tested this by means of a leave-one-out analysis. The combined model was truncated at 95% variance as in the previous experiments.

The results given in Tables 7.9 and 7.10 show – in comparison with Tables 7.7 and 7.8 – a modest increase in performance w.r.t. pt.pt. distance. This encouraging result suggests that parallel analysis provides both faster *and* more general AAMs in a non-parametric fashion. Consequently, simple variance thresholding seems to over-fit the training data slightly by including too many modes of variation.

Table 7.9: PA segmentation results – large models, no subsampling.

Model type	Learning method	Mean pt.pt.	Mean pt.crv.
Grey-scale	Jacobian	5.70±1.18	3.02±0.76
Colour	Jacobian	5.54±1.17	2.93±0.78

Table 7.10: PA segmentation results – small models, no subsampling.

Model type	Learning method	Mean pt.pt.	Mean pt.crv.
Grey-scale	Jacobian	2.87±0.62	1.50±0.39
Colour	Jacobian	2.79±0.60	1.47±0.40

7.7.2 Cardiac Magnetic Resonance Images

For the second case study, short-axis, end-diastolic cardiac MRIs were selected from 14 subjects. MRIs were acquired using a whole-body MR unit (Siemens Impact) operating at 1.0 Tesla. The chosen MRI slice position represented low morphologic complexity, judged by the presence of papillary muscles. Images were acquired using an ECG-triggered breath-hold fast low angle shot (FLASH) cinematographic pulse sequence. Slice thickness=10 mm; field of view=263×350 mm; matrix 256×256. The endocardial and epicardial contours of the left ventricle were annotated manually by placing 33 landmarks along each contour, see Figure 7.13. To fix rotation around the left ventricular (LV) long-axis, the right ventricle (RV) was annotated using 12 landmarks. Prior to further processing, subject 11 was removed due to a scanner calibration error.

Figure 7.13: Example annotation of the left ventricle using 66 landmarks.

Shape, Texture and Combined Model

Figure 7.14 shows all cardiac shapes unaligned (left) and Procrustes aligned (right). Combining the cardiac shape and texture model yields the first three modes of combined variation shown in Figure 7.15. These three modes account for 22%, 20%, 14% of the total variance, respectively.

Segmentation Accuracy

To assess segmentation accuracy an evaluation methodology similar to that of the face case was adopted. The models stemming from the leave-one-out experiments

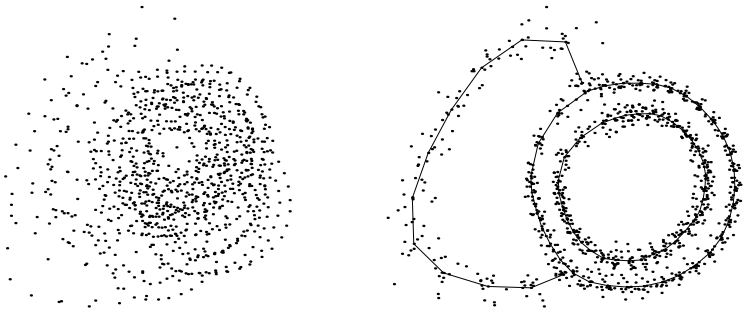


Figure 7.14: Unaligned shapes (left). Aligned shapes (right).

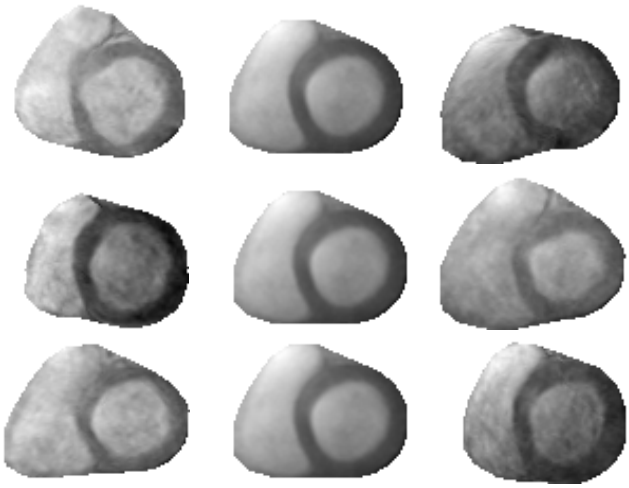


Figure 7.15: Combined deformation using the three largest principal modes (row-wise, top-down).

contained 4192 texture samples on average. Variance-based truncation at a 95% level, yielded 9 combined parameters, whereas parallel analysis settled on 5 parameters on average. No subsampling was used during the training phase. The corresponding results are summarised in Table 7.11 and Table 7.12 using both learning methods. Consistent with the face case we observe that i) parallel analysis provided slightly better accuracy using fewer model modes and ii) no significant difference between the two learning methods existed. The average segmentation time was 29 ms.

Table 7.11: Segmentation results – variance truncation.

Learning method	Mean pt.pt.	Mean pt.crv.
Regression	3.27±1.06	1.38±0.34
Jacobian	3.20±0.91	1.28±0.25

Table 7.12: Segmentation results – parallel analysis.

Learning method	Mean pt.pt.	Mean pt.crv.
Regression	3.02±0.83	1.38±0.31
Jacobian	3.11±0.91	1.36±0.26

FAME also provides a generic method for unsupervised model initialisation. Basically, this is a semi-brute force model-based search in selected model and pose parameters, which exploits the convergence radius of each parameter. To add robustness this is implemented in a candidate scheme, where several initialisation candidates compete to become the initial configuration. For details see [205].

This method has been employed in leave-one-out experiments similar to the above. The results given in Table 7.13 show good correspondence with Table 7.11, suggesting that unsupervised segmentation is plausible without employing a dedicated initialisation method.

Table 7.13: Segmentation using automatic model initialisation (using variance truncation).

Learning method	Mean pt.pt.	Mean pt.crv.
Regression	3.20±1.18	1.31±0.30
Jacobian	3.21±1.02	1.28±0.24

Finally, the alleged robustness to image noise was tested by adding Gaussian noise at different levels, $\sigma^2 \in [0; 40^2]$, to the training set. Image intensities were in the range [0;255]. For each level a leave-one-out study similar to the above was conducted. To improve the robustness we have used the multi-scale feature that FAME provides. Here optimisation is started at the smallest level of a pyramidal set of AAMs and then propagated down towards the original resolution. The results given in Figure 7.16 show a modest 15% increase in pt.crv. accuracy when comparing the original images to the most noise contaminated. Further, we observe that the accuracy degrades gracefully with increasing noise amplitude. Figure 7.17 shows an average segmentation result with $\sigma = 40$.

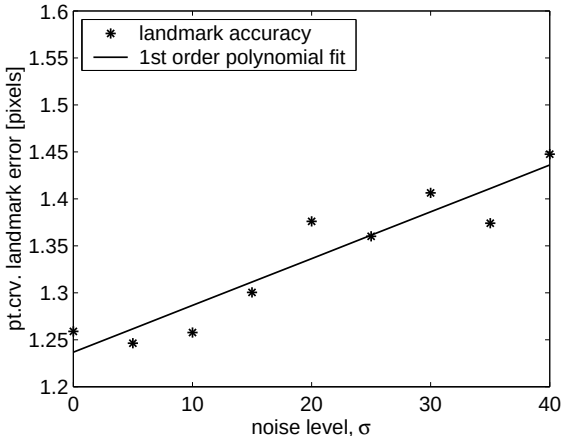


Figure 7.16: AAM segmentation accuracy in the presence of noise.

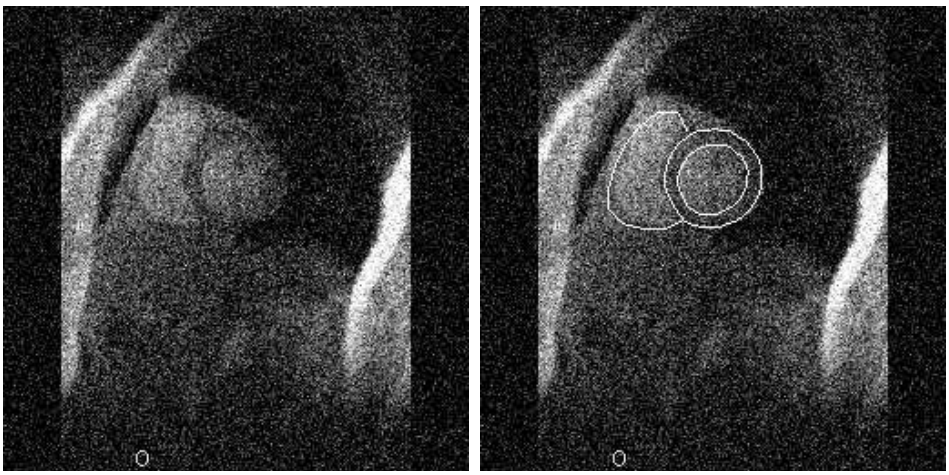


Figure 7.17: Cardiac MRI with added noise, $\sigma = 40$ (left). Corresponding segmentation result.

This concludes the study of cross-sectional cardiac MRI. Here we have treated segmentation of both LV and RV. We note that typically only the segmentation of LV is of interest in, e.g. an ejection fraction study. In this case a hierarchical approach [28] should be used to increase segmentation accuracy. Here, an LV model should be de-coupled from the converged RV-LV model and iterated further, thus relaxing model constraints. See also [212] on this matter.

7.8 Discussion and Conclusion

In this paper we have described an appearance-based software environment for interpretation of images. This is the Flexible Appearance Modelling Environment (FAME), which is based on the well-known Active Appearance Model (AAM) framework. We have summarised the wide range of medical applications that frameworks similar to ours have been applied to. In particular, we have treated two crucial parts in the building process of an AAM. Namely, how to learn the update step used during model fitting, and how to choose the number of model parameters. These issues along with several others have been evaluated in two different case studies upon short-axis cardiac magnetic resonance images and face images. The proposed unsupervised method for choosing the number of model modes proved preferable in all cases. Further, it has been verified that the Jacobian learning scheme is preferable to the regression approach taken in the original AAM formulation. In all studies the AAM variation implemented in FAME has proven accurate, fast and very robust to image noise. An example of how a strong prior redeems an otherwise ill-posed segmentation problem was given in the Cardiac MRI study. This showed a very modest penalty in segmentation accuracy even for very noisy data when using a multi-scale approach.

The performance provided by the framework renders it feasible to interactive analysis of medical images. Further, it allows the usage of general-purpose optimisers to ”fine-polish” the converged search results in a reasonable amount of time [212, 214]. For completely unsupervised image analysis, FAME also provides means of automatic model segmentation. Due to the modular structure of FAME, it is straightforward to extend it to cope with other medical data than 2D images, e.g. time-series of 2D images, coupled images (e.g. from different imaging modalities) etc.

The FAME package including source code and all data material used in this paper has been placed in the public domain. Thus we encourage researchers to apply, extend and embed FAME in future studies both medical and non-medical. Further, we advocate for a general trend between researchers to benchmark their methods against reference data sets and reference implementations, thus providing a much more clear view on the progress in the field of medical image analysis.

Acknowledgements

The following are gratefully acknowledged for their help in this work. Cardiac MRIs were provided by M.D., Ph.D., J. C. Nilsson and M.D., Ph.D., B. A. Grønning, Danish Research Centre for Magnetic Resonance, DRCMR, Hvidovre University Hospital. M. M. Nordstrøm, M. Larsen and J. Sierakowski collected and annotated the face images. We would further like to thank the anonymous reviewers for their helpful comments on an earlier version of this manuscript.

7.A Notes on the FAME Implementation

Texture Synthesis

As shown above, texture synthesis is carried out by

$$\mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_t \mathbf{\Phi}_{c,t} \mathbf{c}. \tag{7.15}$$

For clarity reasons this expression is often written in the AAM literature as

$$\mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi} \mathbf{c}. \tag{7.16}$$

Let,

$$k_t = \mathrm{rank}(\mathbf{\Phi}_t) \quad , \quad k_c = \mathrm{rank}(\mathbf{\Phi}_{c,t}) \tag{7.17}$$

Though it seems tempting to pre-calculate the constant matrix product in (7.15), this should only be done when $M(k_c - k_t) < k_t k_c$. Otherwise, synthesis should be carried out as

$$\mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_t (\mathbf{\Phi}_{c,t} \mathbf{c}). \tag{7.18}$$

To illustrate this, consider the AAM described in [3] having 66 shape parameters and 10 texture parameters and 5040 texture samples. If we assume that these are only moderately correlated, 74 combined parameters would seem plausible. The extra cost of using (7.16) compared to (7.15) is $M(k_c - k_t) - k_t k_c$. Consequently, this amounts to 321820 extra multiplications for each texture synthesis.

Texture Normalisation

Contrary to e.g. [54, 63, 66] FAME does not include a model for scaling and offset of textures. During analysis texture vectors are standardised to zero mean and fixed variance. During synthesis, model textures are fitted to image textures in a least squares sense.

Pose parameterisation

To estimate the Jacobian matrix, pose parameters need to be independent. To obtain equilibrium at zero we use the following representation of pose

$$\rho = [\ (s-1)\ \theta\ t_x\ t_y\ ]^{\mathsf{T}}, \tag{7.19}$$

where s denotes scale, θ orientation and t_x, t_y in-plane translation.

7.B FAME Model Options

This section enumerates the most important model design options offered by FAME.

- *Model reduction* Determines the image reduction prior to model building, i.e. restricts the number of texture samples in the model.
- *Convex hull* Determines whether i) all texture samples inside the convex hull of the mean shape should be modelled or ii) only the texture samples inside closed contours.
- *Tangent space projection* Determines if a linearisation of the shape manifold around the mean pole should be carried out.
- *Learning method* Selects either Principal Component Regression or Jacobian Estimation to build parameter update matrices.
- *Training set subsampling* Option that chooses every n-th example in the training set to build parameter update matrices upon.
- *Model truncation* Controls the number of parameters in the shape, texture and combined models. Either i) by a variance threshold, or ii) using parallel analysis to estimate the number of modes to retain.
- *Warping method* Selects i) software warping, ii) hardware warping, or iii) the benchmarking mode, where the model building process is replaced by a performance test of the two warping methods.

7.C Hardware-assisted AAM Warping

Contemporary graphics hardware is highly optimised for performing piece-wise affine warps. Below we will demonstrate how to exploit this in an AAM framework.

In AAMs two types of warps are carried out. One during the analysis phase (i.e. model building and model optimisation) and a one during synthesis. In the analysis phase many-to-one warps are carried out. Different configurations of the shape model are all warped to same shape-free reference frame, which is typically the mean shape sized to mean size. During synthesis shape-free textures are warped to specific shape configurations.

We approach the graphics hardware through the industry standard for graphics programming: OpenGL. Here a triangular mesh can be defined and images can be projected onto this surface by means of texture mapping. In addition, recent extensions such as hardware-accelerated off-screen buffers, non-dyadic textures etc. can be utilised. For details on the implementation and a discussion of its bottlenecks, refer to [206].

7.D Details on Machinery and Computation Time

To give an impression of the applicability of these models timings were acquired for the face model building phase and the optimisation phase. These are given as reference numbers. For maximal speed a multi-resolution model should be used. The build timings were calculated for both large and small models in colour and grey-scale built on all 37 training examples using the subsampling scheme mentioned in Section 7.4. All timings in Table 7.15 were obtained on similar models but using Jacobian and leave-one-out on the training examples.

All results shown above including the performance measures in Table 7.14 and 7.15 were obtained on an Athlon 1200 MHz equipped with 768 MB RAM using software warping. Except for the last row in Table 7.14 that used a GeForce 2 MX for hardware warping. From Tables 7.14 and 7.15 we can observe that warping is not the bottleneck in the current implementation.

In Table 7.16 we used a Pentium 4 mobile 1700 Mhz equipped with an NVidia GeForce 4 Go graphics board. We have used high performance timers and several hundred repetitions in order to measure accurately in the millisecond range. From this table we can observe that the specific graphics board (and its driver) has a great impact on the performance. Further, since the hardware AAM warping is very heavy on bus i/o the infrastructure between GPU and CPU is of utmost importance.

Table 7.14: Model building and warping (small/large models).

	Gray-scale	Colour
Model build, regression	0:31 / 1:11 mins	1:03 / 2:51 mins
Model build, Jacobian	0:09 / 0:39 mins	0:27 / 1:54 mins
Software analysis warp	1 / 9 ms	3 / 12 ms
Hardware analysis warp	2 / 5 ms	2 / 7 ms

Table 7.15: Optimisation timings (small/large models).

	Gray-scale	Colour
Mean number of iterations	11.4 / 11.3	8.7 / 8.2
Mean optimisation time	129 / 581 ms	281 / 1253 ms
1 optimisation iteration	11 / 51 ms	32 / 153 ms

Table 7.16: Warp timings – GeForce 4 Go (large model).

	Gray-scale	Colour
Software analysis warp	5.4 ms	7.4 ms
Hardware analysis warp	1.5 ms	2.4 ms

CHAPTER 8

Wavelet Enhanced Appearance Modelling

Mikkel B. Stegmann, Søren Forchhammer, and Timothy F. Cootes

Abstract

Generative segmentation methods such as the Active Appearance Models (AAM) establish dense correspondences by modelling variation of shape and pixel intensities. Alas, for 3D and high-resolution 2D images typical in medical imaging, this approach is rendered infeasible due to excessive storage and computational requirements. This paper extends the previous work of Wolstenholme and Taylor where Haar wavelet coefficient subsets were modelled rather than pixel intensities. In addition to a detailed review of the method and a discussion of the integration into an AAM-framework, we demonstrate that the more recent bi-orthogonal CDF 9-7 wavelet offers advantages over the traditional Haar wavelet in terms of synthesis quality and accuracy. Further, we demonstrate that the inherent frequency separation in wavelets allows for simple band-pass filtering, e.g. edge-emphasis. Experiments using Haar and CDF 9-7 wavelets on face images have shown that segmentation accuracy degrades gracefully with increasing compression ratio. Further, a proposed weighting scheme emphasizing edges was shown to be significantly more accurate at compression ratio 1:1, than a conventional AAM. At higher compression ratios the scheme offered both a decrease in complexity *and* an increase in segmentation accuracy.

Keywords: registration, compression, atlases, deformable models, active appearance models, wavelets, face images

8.1 Introduction

Generative models capable of synthesising photo-realistic images of objects have been shown to be powerful tools for image interpretation. Explicitly modelling the value of every pixel covering an object is feasible for low-resolution 2D images. However, for high-resolution 2D images, 2D and even 3D time-series, this approach is rendered infeasible due to excessive storage and computational requirements. This is especially unfortunate due to the many information-rich medical imaging modalities emerging and maturing these days. We therefore address this problem by applying wavelet compression to the widely used generative model; the Active Appearance Model (AAM) [52, 88]. Refer to [65, 213] for reviews of various applications.

Since wavelets provide an approximate decorrelated representation of each training image, we aim at choosing consistent subsets of wavelet coefficients over the complete training set. The AAM is subsequently built using these subsets in order to lessen the computational and storage requirements. The major challenge in this context is to choose an appropriate wavelet basis and a subset that affects the final performance measure as little as possible. In this study we focus on the segmentation capabilities of AAMs.

The presented work is validated on an example of very complex biological variation in both shape and appearance. It consists of a cohort of face images. Examples of medical use of statistical face models include the work of Hutton et al. [125], which aims at a better understanding of syndromic facial dysmorphologies in a study of the Noonan syndrome.

8.2 Related Work

Model size concerns have been present since the introduction of AAMs [52, 88]. The initial models contained $\sim$10.000 pixels and were built on sub-sampled input images of approximately 200×300 pixels [65]. Aiming at reducing computational costs Cootes et al. [51] used a sub-sampling scheme to reduce the texture model by a ratio of 1:4. The scheme selected a subset based on the ability of each pixel to predict corrections of the model parameters. When exploring different multi-band appearance representations Stegmann and Larsen [221] studied the segmentation accuracy of face AAMs at different scales in the range 850 – 92.118 pixels obtained by pixel averaging. Mitchell et al. [161] introduced 3D AAMs and noticed that a current limitation was the requirement of heavy down-scaling of their input images (3D cardiac MRI and 2D+time ultra sound).

This paper extends directly on the original work of Wolstenholme and Taylor [249, 250], which introduced the idea of modelling wavelet coefficient subsets in Active Appearance Models. They augmented an AAM with the Haar wavelet and carried

out a set of experiments on brain MRI using a fixed model size compression ratio of 1:20 using both a multi-resolution and single resolution model. Other compression ratios were mentioned, however, no qualitative results were given for these.

8.3 Active Appearance Models

Active Appearance Models are generative models capable of synthesising images of a given object class. By estimating a compact and specific basis from a training set, model parameters can be adjusted to fit unseen images and hence perform image interpretation. The modelled object properties are usually shape and pixel intensities (here denoted *texture*). Training objects are defined by marking up each example image with points of correspondence. Using prior knowledge of the optimisation space, AAMs can be rapidly fitted to unseen images, given a reasonable initialisation.

Variability is modelled by means of principal component analyses (PCA). Prior to PCA modelling shapes are Procrustes aligned and textures are warped into a shape-free reference frame and sampled. Let there be given P training examples for an object class, and let each example be represented by a set of N landmark points and M texture samples. Let $\mathbf{s}$ and $\mathbf{t}$ denote a synthesised shape and texture and let $\bar{\mathbf{s}}$ and $\bar{\mathbf{t}}$ denote the corresponding means. Synthetic examples are parameterised by $\mathbf{c}$ and generated by

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{\Phi}_{s,c}\mathbf{c} \quad , \quad \mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_{t,c}\mathbf{c} \tag{8.1}$$

where $\mathbf{\Phi}_{s,c}$ and $\mathbf{\Phi}_{t,c}$ are composed of eigenvectors estimated from the training set dispersions. An object instance, $\{\mathbf{s}, \mathbf{t}\}$, is synthesised into an image by warping the pixel intensities of $\mathbf{t}$ into the geometry of the shape $\mathbf{s}$.

An AAM is matched to an unseen image using a least-squares criterion by an iterative updating scheme. This is based on a fixed Jacobian estimate [54], or, originally, a principal component regression [52]. For this work we have chosen the former approach in a setup treated in detail elsewhere [213]. For further details on AAMs refer to [52, 54, 65].

8.4 Wavelets

Wavelets are a family of basis functions that decompose signals into both space and frequency. In the following we use the discrete wavelet transform (DWT), which can be viewed as a set of linear, rank-preserving matrix operations. In practice these are carried out in a convolution scheme known as the fast wavelet transform (FWT) [5] where an image is decomposed by a high-pass filter into a set of detail wavelet sub bands, and by a low-pass filter into a scaling sub band. These bands are then

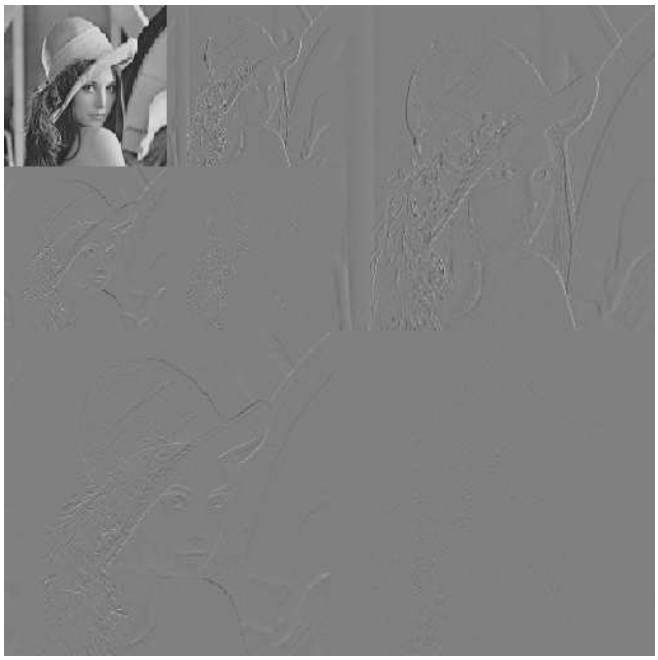


Figure 8.1: The wavelet coefficients of a two-level octave decomposition using the Haar wavelet.

down-sampled and can be further decomposed. We use the dyadic (octave) decomposition scheme that successively decomposes the scaling sub band, yielding a discrete frequency decomposition. Alternative decomposition schemes include the wavelet packet basis where successive decompositions are carried out in the detail sub bands as well.

Figure 8.1 shows a two-level octave wavelet decomposition. The first, third and fourth quadrants are the detail sub bands and stem from the initial decomposition (level 1). The four sub-quadrants of the second quadrant stem from the second decomposition (level 2) with the scaling sub band at the top left corner.

Wavelets are invertible which is typically achieved by orthonormality. Wavelet transforms can thus be considered a rotation in function space, which in addition – through successive decompositions – adds a notion of scale. This scale property lends itself nicely to progressive signal processing. Wavelets for image compression are designed to perform rotations that decorrelate image data by using vanishing moments. Wavelet coefficients close to zero can thus be removed with minimal impact on the reconstruction.

Bi-orthogonal wavelets will also be considered in the following. These are not strictly orthogonal and therefore have odd finite length and linear phase. They come in pairs of analysis and synthesis filters, together forming a unitary operation.

8.5 Wavelet Enhanced Appearance Modelling

This section introduces a notation for wavelet compression and describes how it can be integrated into an AAM framework thereby obtaining a Wavelet enHanced Appearance Model (WHAM).

First, let an n-level wavelet transform be denoted by

$$\mathcal{W}(\mathbf{t}) = \mathbf{\Gamma t} = \hat{\mathbf{w}} = [\ \hat{\mathbf{a}}^\mathsf{T}\ \hat{\mathbf{u}}_1^\mathsf{T}\ \cdots\ \hat{\mathbf{u}}_n^\mathsf{T}\ ]^\mathsf{T} \tag{8.2}$$

where $\hat{\mathbf{a}}$ and $\hat{\mathbf{u}}$ denote scaling and detail wavelet coefficients, respectively. For 2D images each set of detail coefficients is an ensemble of horizontal, vertical and diagonal filter responses. Compression is now obtained by a truncation of the wavelet coefficients

$$\mathcal{C}(\hat{\mathbf{w}}) = \mathbf{C}\hat{\mathbf{w}} = \mathbf{w} = [\ \mathbf{a}^\mathsf{T}\ \mathbf{u}_1^\mathsf{T}\ \cdots\ \mathbf{u}_n^\mathsf{T}\ ]^\mathsf{T} \tag{8.3}$$

where $\mathbf{C}$ is a modified identity matrix, with rows corresponding to truncated coefficients removed.

As in [250] a WHAM is built on the truncated wavelet basis, $\mathbf{w} = \mathcal{C}(\mathcal{W}(\mathbf{t}))$, rather than the raw image intensities in $\mathbf{t}$. This splits all texture-related matrices into scale-portions. For the texture PCA of wavelet coefficients we have

$$\mathbf{w} = \overline{\mathbf{w}} + \mathbf{\Phi}_w \mathbf{b}_w \quad \Leftrightarrow \tag{8.4}$$

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{a}} \\ \overline{\mathbf{u}}_1 \\ \vdots \\ \overline{\mathbf{u}}_n \end{bmatrix} + \begin{bmatrix} \mathbf{\Phi}_a \\ \mathbf{\Phi}_{u_1} \\ \vdots \\ \mathbf{\Phi}_{u_n} \end{bmatrix} \mathbf{b}_w$$

where $\mathbf{\Phi}_w$ is the eigenvectors of the wavelet coefficient covariance matrix. Rearranging this into scaling and detail terms we get

$$\mathbf{a} = \overline{\mathbf{a}} + \mathbf{\Phi}_a \mathbf{b}_w \tag{8.5}$$

and

$$\{\mathbf{u}_i = \overline{\mathbf{u}}_i + \mathbf{\Phi}_{u_i} \mathbf{b}_w\}_{i=1}^n. \tag{8.6}$$

The texture model is thus inherently multi-scale and may be used for analysis/synthesis at any given scale. Motivations for doing so include robustness and computational efficiency. Compared to multi-scale AAMs this also gives a major decrease in storage requirements. However, since the PCA is calculated at full scale (including all detail bands) we can not expect this to be equal to the separate scale-PCAs of conventional multi-scale AAMs.

Using non-truncated orthogonal wavelets (i.e. $\mathbf{C} = \mathbf{I}$), $\mathcal{W}(\mathbf{t})$ is a true rotation in texture hyperspace. Hence the wavelet PCA is a rotation of the original intensity PCA, i.e. $\mathbf{\Phi}_w = \mathbf{\Gamma}\mathbf{\Phi}_t$, iif $\mathcal{W}$ is fixed over the training set. The PC scores are identical,

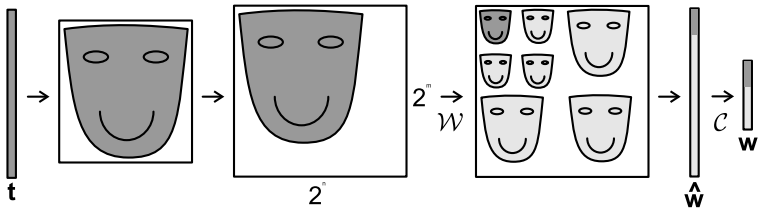


Figure 8.2: Two-level wavelet decomposition of a texture vector.

$\mathbf{b}_w = \mathbf{b}_t$. If $\mathbf{C}$ is chosen to truncate the wavelet basis along directions with near-zero magnitude, wavelet PC scores obviously resembles the original PC scores closely.

Direct usage of the sparse matrix $\mathbf{\Gamma}$ is excessively slow. Instead the fast wavelet transform (FWT) is applied. Figure 8.2 shows the stages of transformation. First, a normalised texture vector is rendered into its equivalent shape-free image. Secondly, the shape-free image is expanded into a dyadic image representation to avoid any constraints on n due to image size. This image is then transformed using FWT and rendered into the vector $\hat{\mathbf{w}}$ by masking out areas outside the octave representation of the reference shape. Eventually, $\hat{\mathbf{w}}$ is truncated into $\mathbf{w}$.

8.5.1 Free Parameters and Boundary Effects

To apply the above, the type of wavelet, $\mathcal{W}$, must be chosen and values for n and $\mathbf{C}$ must be determined. Being a key issue, the estimation of $\mathbf{C}$ is described in a later section.

The choice of wavelet type depends on two factors, i) the nature of the data, and ii) computational issues. Data containing sharp edges suggests sharp filters such as the Haar wavelet and smooth data requires smooth wavelets. However, the price comes in form of increased computational load induced by the extra filter-taps. Since the wavelets operate in a finite discrete domain, traditional boundary issues arise. Consequently, the longer, smoother wavelets suffer the most. To calculate filter responses for a full image, boundaries are typically extended by mirroring. The width of the boundary extension is half the width of the wavelet filter. For even-length wavelets such as Haar the symmetry point is placed between the edge pixel and the first boundary extension pixel. For odd-length wavelets the symmetry point is placed on the edge pixel. Normally this is carried out as a rectangular extension but in this application the extension should adapt to the shape of the texture image.

When using an out-of-the-box wavelet package for rectangular domains, a poor mans solution is to approximate this with a shape adaptive horizontal/vertical mirroring of the dyadic shape image. The width of the mirroring should be $<\text{filter width}>\times 2^{n-1}$.

8.5.2 WHAM Building

Building a WHAM can be summarised in five major steps:

1. Sample all training textures into $\{\mathbf{t}_i\}_{i=1}^{P}$.
2. Transform $\{\mathbf{t}_i\}_{i=1}^{P}$ into $\{\hat{\mathbf{w}}_i\}_{i=1}^{P}$.
3. Estimate $\overline{\hat{\mathbf{w}}}$ and $\mathbf{C}$.
4. Truncate $\{\hat{\mathbf{w}}_i\}_{i=1}^{P}$ into $\{\mathbf{w}_i\}_{i=1}^{P}$.
5. Build an AAM on $\{\mathbf{w}_i\}_{i=1}^{P}$.

Further, all incoming textures in subsequent optimisation stages should be replaced with their truncated wavelet equivalents, i.e. $\mathcal{C}(\mathcal{W}(\mathbf{t}))$. The synthesis of a WHAM is the reverse of Figure 8.2, again with appropriate masking and mirroring. Truncated wavelet coefficients are reconstructed using the corresponding mean values:

$$\hat{\mathbf{w}}_{\text{synth}} = \mathcal{C}_{\text{synth}}(\mathbf{w}) = \mathbf{C}^{\mathsf{T}}\mathbf{w} + \mathbf{C}_{\text{synth}}\overline{\hat{\mathbf{w}}} \tag{8.7}$$

where $\mathbf{C}_{\text{synth}}$ is a modified identity matrix, with rows corresponding to non-truncated coefficients zeroed.

8.5.3 A Note on Representation

Depending on the AAM implementation, the conversion from $\mathbf{t}$ to the shape-free image might be avoided since this is often the canonical representation of $\mathbf{t}$ after warping. The calculation of the truncation can also be done directly in image-space to avoid unnecessary transforms from vector to image-space, et cetera.

8.5.4 Wavelet Coefficient Selection

For segmentation purposes the optimal $\mathbf{C}$ is given by the minimum average error between the optimised model shape and the ground truth shape over a test set of t examples

$$\arg\min_{\mathbf{C}} \left(\sum_{i=1}^{t} |\mathbf{s}_{i,\text{model}} - \mathbf{s}_{i,\text{ground truth}}|^2 \right), \tag{8.8}$$

subject to the constraint that $\mathbf{C}$ has m rows. This gives rise to a compression of ratio $1 : M/m$.

Contrary to traditional image compression, quantization and entropy coding are not carried out. Consequently, this compression ratio does not commensurate to those in the compression literature.

Unfortunately, direct optimisation of Equation 8.8 is not feasible since each cost function evaluation involves building a complete model from the training set and a subsequent evaluation on a test set. Alternatively one can design $\mathbf{C}$ based on prior beliefs.

The traditional approach when dealing with compression of training data ensembles, also taken by [250], is to let $\mathbf{C}$ preserve per-pixel variance over the training set. This is accomplished by constructing $\mathbf{C}$ to select the m largest coefficients from

$$\kappa = \sum_{i=1}^{P} (\hat{\mathbf{w}}_i - \overline{\hat{\mathbf{w}}}) \odot (\hat{\mathbf{w}}_i - \overline{\hat{\mathbf{w}}}), \tag{8.9}$$

where $\odot$ denotes the Hadamard product (i.e. element-wise multiplication). We further assume spatial coherence and regularise this scheme prior to coefficient selection by a smoothing of κ (in the image domain) using a convolution with a suitable kernel.

In summary, this prior equates variance with interesting signal.

8.5.5 Signal Manipulation

An important added bonus from the wavelet representation is that certain signal manipulations become exceedingly simple. Due to the band separation, frequency response modifications are easily carried out. The simplest modification is to change the norm of the high- and low-pass filters used. To obtain orthonormality the filters must be normalised. With a minor abuse of notation this is denoted by $||\mathcal{W}||_2 = 1$. We call this the *normalised case*. To emphasise high-frequency content $||\mathcal{W}||_2$ must be less than one. In the following we propose to use the norm $||\mathcal{W}||_2 = 1/\sqrt{2}$ and call this the *weighted case*.

8.5.6 Extension to Multi-channel AAMs

Support for multi-channel images such as RGB [87], Edge AAMs [64] or composite representations [221] can be implemented in different ways. A simple approach is to wavelet transform each texture band separately with non-coupled truncation schemes. However, this often leads to situations where one multi-band pixel is partially truncated and partially included into the model. If this is not desirable, a composite measure must be used when estimating $\mathbf{C}$, e.g. the sum of band variances at each pixel position. If pixels are correlated across bands as in e.g. RGB images, the wavelet decomposition can be carried out as a 3D wavelet transformation, resulting in higher compression (or better quality).

8.6 Experimental Results

To assess the segmentation accuracy of a WHAM a set of experiments was carried out on 37 still images of people facing the camera acquired as 640×480 pixel, 8 bit, grey-scale images and annotated using 58 landmarks, see Figure 8.3 (left).

The average landmark distance from model to ground truth (pt.pt.) was used as performance measure. Model searches were initialised by displacing the mean configuration $\pm 10\%$ of its width and height in x and y from the optimal position. To reduce statistical fluctuation and reduce pessimistic bias leave-one-out evaluations were carried out. Two wavelet bases were evaluated; the orthogonal Haar wavelet (2 filter taps) and the widely used bi-orthogonal CDF 9-7 wavelet (9 and 7 filter taps). Both of these were tested in the normalised and the weighted case using compression ratios in the range 1:1 – 1:40 and compared to a standard AAM. This lead to a total of 37×2×2×9+37×1 = 1369 models. All experiments used three decomposition levels. Results are shown in Figure 8.4 and 8.5. The standard AAM contained 31224 pixels on average over all leave-one-out experiments.

In the normalised case we observe that the accuracy degrades gracefully with increasing compression ratio. Though consistently better on average at all ratios, the CDF 9-7 wavelet is not markedly better than the simple Haar wavelet in terms of accuracy. In the weighted case we observe a considerable improvement in accuracy for almost all compression ratios. The non-overlapping notches in the box plots in Figure 8.5 show that the median of the standard AAM is significantly worse than both the median of the uncompressed Haar and the CDF 9-7 wavelet at a 0.05 level.

To investigate this behaviour further, Figure 8.6 shows the selected wavelet coefficients using the CDF 9-7 wavelet at ratio 1:10 using both normalised and weighted filters. The normalised case tends to preserve mostly low-frequency content, while the weighted case distributes wavelets coefficients near image details more evenly

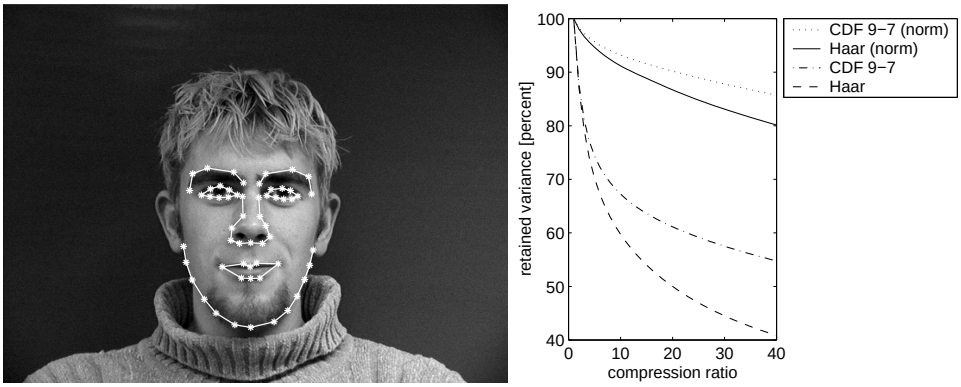


Figure 8.3: Left: Example annotation of a face using 58 landmarks. Right: Retained wavelet coefficient variance over the training set as a function of factor by which the number of wavelet coefficients have been reduced.

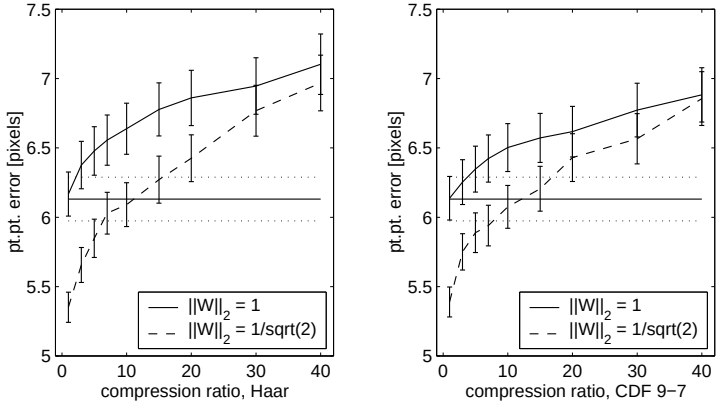


Figure 8.4: Average segmentation error vs. compression ratio. Error bars are one standard error. Horizontal lines show values for a standard AAM.

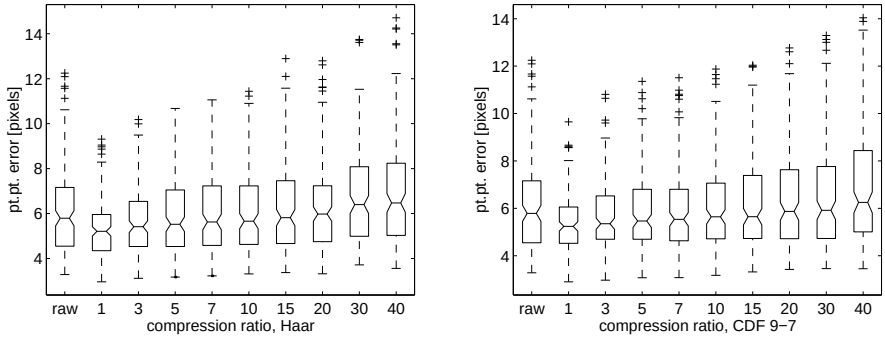


Figure 8.5: Boxplots of segmentation error in the weighted case vs. compression ratio including a standard AAM (raw). Whiskers are 1.5 IQR at maximum.

across levels. Together with the nature of the weighting this leads to models with more emphasis on edges.

The amount of retained wavelet coefficient variance over the training set is plotted in Figure 8.3 (right). As expected, the CDF 9-7 performs consistently better than the Haar wavelet. Further, the normalised wavelets are seen to put relatively more variance into fewer components, when compared to the weighted.

Finally, Figure 8.7 shows the first combined mode of texture and shape deformation for an uncompressed model and two compressed models at compression ratio 1:10. Subtle blocking artefacts from the Haar wavelet are present, while the smooth 9-7 leaves a more visually pleasing synthesis result. Quantitatively, the 9-7 wavelet also offered a better reconstruction of the training textures in terms of mean squared error (MSE).

Figure 8.6: Selected wavelet coefficients for the face training set (CDF, ratio 1:10). The scaling coefficients are shown in the upper left. Left: $||\mathcal{W}||_2 = 1$. Right: $||\mathcal{W}||_2 = 1/\sqrt{2}$.

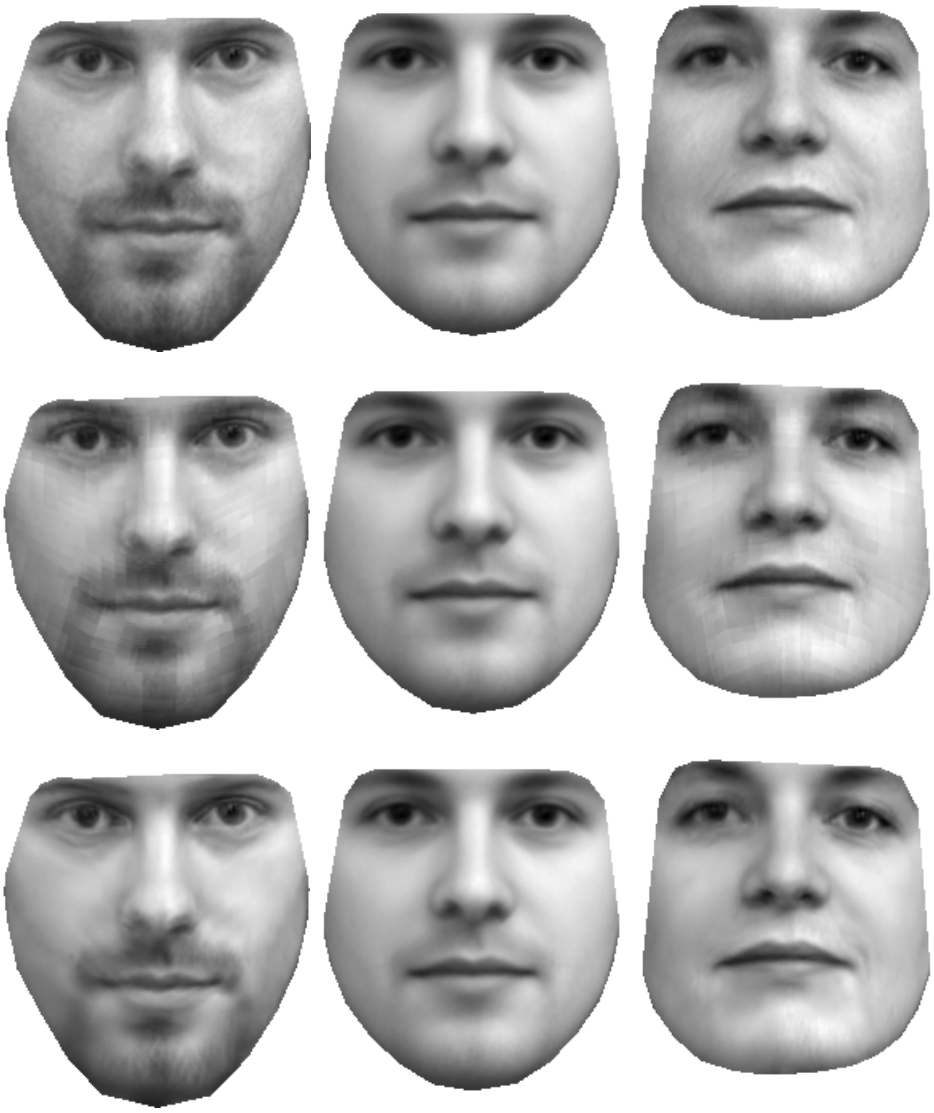


Figure 8.7: The first combined mode of texture and shape variation; $c_1 = \{-3\sigma_1,\ 0,\ 3\sigma_1\}$. Top: AAM (i.e. uncompressed). Middle: WHAM (Haar, ratio 1:10). Bottom: WHAM (CDF 9-7, ratio 1:10).

8.7 Discussion

Although no significant difference was observed between the two wavelets in terms of segmentation accuracy; the CDF 9-7 wavelet performed consistently better. Further, the CDF 9-7 also provided the best synthesis quality. However, if low computational complexity is required the Haar wavelet should be chosen.

Our case study supports the expected behaviour that it is the high-frequency image content that provides the segmentation accuracy. Variance preserving filters are optimal for reconstruction but not necessarily for segmentation. Inherent scale-separation has been obtained in the presented unified AAM and DWT framework and enabled a very simple method for favouring high-frequency content. This is obtained while still retaining the robustness gained from a low frequency representation of the object.

This paper extends the previous work of Wolstenholme and Taylor where an AAM was augmented with the Haar wavelet and evaluated on brain MRI using a fixed compression ratio of 1:20. Our work validates their previous findings in a different case study using a more thorough evaluation methodology. In addition, the more recent CDF 9-7 wavelet is evaluated and compared to the Haar wavelet. Finally, we have demonstrated that the inherent frequency separation in wavelets allows for simple band-pass filtering, which enables compression schemes that both decrease complexity *and* increase segmentation accuracy.

8.8 Future Work

Future work should include investigation of more elaborate coefficient selection and sub band weighting schemes and the (straightforward) extension to n-D images. The latter could be applied to medical data such as 3D and 4D cardiac MRI, 3D brain MRI, ultra sound etc. Exploiting the inherent scale properties of a WHAM is also attractive in order to obtain computationally cheaper and more robust means of optimisation. Finally, earlier work on denoising in the wavelet domain could also be incorporated into this framework.

8.9 Conclusion

This work has described how an AAM framework can be augmented with wavelet compression to reduce model complexity to successfully cope with large medical data sets. Experimental validation has been carried out using two different wavelet bases at nine different compression ratios. It was found that segmentation accuracy degrades gracefully with increasing compression ratio. Our studies indicated that the CDF 9-7 wavelet should be considered as a worthy alternative to the

Haar wavelet. Further, it was shown that a simple weighting scheme could enable wavelets to both decrease complexity *and* increase segmentation accuracy.

We anticipate that wavelet-based appearance modelling will become an important technique with many applications, in particular within medical image interpretation.

CHAPTER 9

Corpus Callosum Analysis using MDL-based Sequential Models of Shape and Appearance

Mikkel B. Stegmann, Rhodri H. Davies, and Charlotte Ryberg

Abstract

This paper describes a method for automatically analysing and segmenting the corpus callosum from magnetic resonance images of the brain based on the widely used Active Appearance Models (AAMs) by Cootes et al. Extensions of the original method, which are designed to improve this specific case are proposed, but all remain applicable to other domain problems. The well-known multi-resolution AAM optimisation is extended to include sequential relaxations on texture resolution, model coverage and model parameter constraints. Fully unsupervised analysis is obtained by exploiting model parameter convergence limits and a maximum likelihood estimate of shape and pose. Further, the important problem of modelling object neighbourhood is addressed. Finally, we describe how correspondence across images is achieved by selecting the minimum description length (MDL) landmarks from a set of training boundaries using the recently proposed method of Davies et al. This MDL-approach ensures a unique parameterisation of corpus callosum contour variation, which is crucial for neurological studies that compare reference areas such as rostrum, splenium, et cetera. We present quantitative and qualitative results that show that the method produces accurate, robust and rapid segmentations in a cross sectional study of 17 subjects, establishing its feasibility as a fully automated clinical tool for analysis and segmentation.

Keywords: registration, atlases, deformable models, active appearance models, MDL landmarking, corpus callosum, brain MRI

9.1 Introduction

Corpus callosum is the nervous tissue that connects the two cerebral hemispheres of the human brain. Many neurological studies indicate that the size and shape of the corpus callosum are related to gender, age, neurodegenerative diseases et cetera [31, 81, 152]. The gold standard for such morphometry studies is magnetic resonance imaging, which allows acquisition of accurate images of the anatomy and function of the human brain. However, obtaining manual tracings of the corpus callosum is both time-consuming, error-prone and operator dependent. Instead, medical image analysis should aim at replacing this task with fully automated and efficient methods eliminating subjectivity. In this paper we demonstrate that the generic Active Appearance Models (AAMs) [52, 88] can be adapted to corpus callosum analysis. The efficacy of this approach will be assessed thorough a detailed quantitative and qualitative validation.

Contrary to manually defined contours, AAMs provide a reference coordinate system for every new example. This makes propagation of reference areas such as the rostrum, genu, truncus, isthmus and splenium a trivial matter. Further, regression analyses upon gender, age, motor abilities, lifestyle et cetera can be carried out directly since shape variation and size are both sensibly and compactly encoded.

This paper is accompanied by a technical report [211] containing background information, preliminary results and elaborations omitted here.

9.2 Related Work

Many attempts on segmenting the corpus callosum have been carried out. However, the majority only gives qualitative results based on one or a few images. Below, recent contributions, which are either related to our approach, give quantitative results, or do both are chronologically summarised. All performance measures mentioned in this section, e.g. pt.pt./pt.crv., are described in Section 9.5.

Staib and Duncan [202] introduced a deformable model, which was based on a Fourier decomposition of object boundaries. Case studies were carried out on synthetic images and two MRI cases, one being a single mid-sagittal corpus callosum image. No quantitative results were given.

Székely et al. [230] developed shape models also based on a Fourier parameterisation and applied these for MRI corpus callosum segmentation in a case study carried out on 30 subjects. The models were initialised manually using an operator-specified AC-PC line as reference. No quantitative results were given.

Cootes et al. [48] proposed to implement deformable anatomical atlases using AAMs in a case study of 72 brain MRIs focusing on the ventricles, the caudate nucleus and the lentiform nucleus. A leave-one-out cross-validation using manual initialisation produced an average accuracy of 2.4 pixels pt.pt. and 1.2 pixels pt.crv.

Lundervold et al. [152] presented a method for MRI corpus callosum segmentation demonstrated on 10 subjects. The method consisted of a combination of a shape model and a multi-spectral tissue classification. The latter was based on three MR pulse sequences; FLASH, DESS, and FISP. Further, thresholding and a morphological cleaning step were needed. The method required no manual interaction and it was concluded that it was promising for clinical use, although no quantitative results were given.

Brejl and Sonka [43] reported a fully automated method for corpus callosum segmentation in MRI. An initial coarse position was determined using a shape-variant Hough transform based on an annotated training set. Refinement was carried out using i) snakes [134] or, ii) a dynamic programming approach, based on a cost function automatically designed from the training set. Trained on 15 subjects and tested on 75 subjects yielded an average segmentation error of 0.7 pixels semi-pt.crv. and 1.2 pixels semi-pt.crv. using either refinement method, respectively.

Van Ginneken et al. [243] augmented Active Shape Models (ASMs) [68] with a non-linear NN-classifier based on sequential forward and backward selection of moments of histograms over multiple scales. The method was evaluated on the data set in [43] using hold-out, i.e. 45 images for training and 45 images for testing. A standard ASM and the non-linear ASM yielded average shape overlaps of 0.62 and 0.81, respectively. Fitting the shape model directly to the ground truth yielded an upper bound of 0.89.

9.3 Data Material

The data material used in this study comprises 17 cross-sectional, mid-sagittal magnetic resonance images (MRI) of the human brain obtained from 17 subjects using a whole body MR scanner (Siemens). MR pulse sequence: MPRAGE; matrix size: 256×256 pixels; subject age: 66/83/77 years (min/max/median).

Corpora callosa were annotated manually by drawing one closed, landmark-free contour in each image. These contours are shown in Figure 9.1, compensated for differences in size and orientation. The corresponding images are shown in [211].

Figure 9.1: Aligned corpus callosum annotations from 17 mid-sagittal MRI, shown row-wise with the Procrustes mean shape in the lower right corner.

9.4 Methods

9.4.1 Active Appearance Models

Active Appearance Models (AAMs) [52, 88] are generative models capable of synthesising images of a given object class. By estimating a compact and specific basis from a training set, model parameters can be adjusted to fit unseen images and hence perform image interpretation. The modelled object properties are shape and pixel intensities (called *texture*). Training objects are defined by marking up each example image with points of correspondence. Variability is modelled by means of principal component analyses (PCA). Prior to PCA modelling shapes are Procrustes aligned and textures are warped into a shape-free reference frame and sampled. Let there be given P training examples, and let each example be represented by a set of N landmark points and M texture samples. Let $\mathbf{s}$ and $\mathbf{t}$ denote a synthesised shape and texture and let $\bar{\mathbf{s}}$ and $\bar{\mathbf{t}}$ denote the corresponding means. New instances can be generated by adjusting the PC scores, $\mathbf{b}_s$ and $\mathbf{b}_t$ in

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{\Phi}_s \mathbf{b}_s \tag{9.1}$$

and

$$\mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_t \mathbf{b}_t, \tag{9.2}$$

where $\mathbf{\Phi}_s$ and $\mathbf{\Phi}_t$ are eigenvectors of the shape and texture dispersions estimated from the training set. To recover any correlation between shape and texture and to obtain a combined parameterisation, $\mathbf{c}$, the values of $\mathbf{b}_s$ and $\mathbf{b}_t$ over the training set are combined in a third PCA,

$$\begin{bmatrix} \mathbf{W}_s \mathbf{\Phi}_s^{\mathrm{T}} (\mathbf{s} - \bar{\mathbf{s}}) \\ \mathbf{\Phi}_t^{\mathrm{T}} (\mathbf{t} - \bar{\mathbf{t}}) \end{bmatrix} = \begin{bmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_t \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_{c,s} \\ \mathbf{\Phi}_{c,t} \end{bmatrix} \mathbf{c} = \mathbf{\Phi}_c \mathbf{c}. \tag{9.3}$$

where $\mathbf{W}_s$ denotes a diagonal matrix weighting pixel distances against intensities. Synthetic examples, parameterised by $\mathbf{c}$, are generated by

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{\Phi}_s \mathbf{W}_s^{-1} \mathbf{\Phi}_{c,s} \mathbf{c} \tag{9.4}$$

and

$$\mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_t \mathbf{\Phi}_{c,t} \mathbf{c} \tag{9.5}$$

and rendered into an image by warping the pixel intensities of $\mathbf{t}$ into the geometry of the shape $\mathbf{s}$.

An AAM is matched to an unseen image using a least-squares criterion by an iterative updating scheme. This is based on a fixed Jacobian estimate [54], or, originally, a principal component regression [52]. For this work we have chosen the former approach, which is treated in further detail elsewhere [213]. For further details on AAMs refer to [52, 54, 65].

9.4.2 Landmark Placement

Establishing points of correspondence, so-called *landmarks*, on training examples by hand can not only be time-consuming, tedious and error-prone, but in some cases directly impossible when no "good" anatomical points are available. In this matter the term good refers to if landmarks can be identified with high degrees of reproducibility and repeatability.

We manually identify two landmarks on the corpus callosum; one at the rostrum and one at the splenium. Further, 78 semi-landmarks were interpolated along the contour using a uniform arc-length parameterisation.

To supplement this, a method for automatic landmarking proposed by Davies et al. [75, 78] was applied. This is a minimum description length (MDL) framework that – in the spirit of Occam's razor – selects the landmarks, which lead to the minimum description length of the Point Distribution Model (PDM) [68] built on the training set. Procrustes alignment was used to obtain an initial alignment and this was optimised using the MDL objective function. Landmarks with a minimal description length were calculated for all shapes and the alignment was inverted so that the MDL-based landmarks were given in image coordinates.

9.4.3 Background Awareness

In medical images, organs and structures typically have a very homogenous appearance. Since AAMs only model the object appearance, one can find that the cost function easily can have a minimum inside the object, where both a plausible shape and good texture fit can be obtained. Unfortunately, corpora callosa are no exception to this. A partial remedy for this is to model the convex hull of the shape points. However, this only helps where the shape is concave. Further, this may lead to inclusion of unwanted texture variation, e.g. the appearance of the fornix brain structure in our case.

Consequently, we propose a simple scheme for modelling background variation by adding normals outwards from the shape, much inspired from the precursor to AAMs; the Active Shape Models [68]. These normals are denoted *whiskers* and added implicitly during texture sampling with a scale relative to the current shape size. Texture samples obtained by sampling along whiskers are concatenated to the texture vector with a fixed and uniform weight compared to the conventional AAM texture samples. This provides a simple weighted method for adding background awareness to an AAM.

9.4.4 Sequential Relaxation of Model Constraints

To improve both convergence and robustness AAM search can be performed at different resolutions. A widely used scheme is to recursively decimate the input image

thereby creating a pyramid representation of images at different scales. This can be carried out on the training set, thus producing one model for each level of the pyramid, see e.g. ref [65]. The details of how this is implemented in the current framework is given and extended below.

Typically, model searches are initialised at the top level and propagated toward the full image resolution. Let $\mathbf{c}_i$ denote the converged model configuration at level i. The corresponding shape, $\mathbf{s}_i$ is propagated to level $i+1=j$ by $\mathbf{s}_j = 2\mathbf{s}_i$ and the texture is sampled into $\mathbf{t}_j$. Assuming that the optimisation at level i produced an actual match, $\mathbf{c}_j$ can now be estimated by projecting $\mathbf{s}_j$ and $\mathbf{t}_j$ into the corresponding combined eigenspace

$$\hat{\mathbf{c}}_j = \mathbf{\Phi}_{c,j}^{\mathsf{T}} \begin{bmatrix} \mathbf{W}_{s,j}\mathbf{\Phi}_{s,j}^{\mathsf{T}}(\mathbf{s}_j - \overline{\mathbf{s}}_j) \\ \mathbf{\Phi}_{t,j}^{\mathsf{T}}(\mathbf{t}_j - \overline{\mathbf{t}}_j) \end{bmatrix} \tag{9.6}$$

and model optimisation can continue at level j. If $\mathbf{\Phi}_{s,i} = \mathbf{\Phi}_{s,j}\ \forall\ i,j$ then $\mathbf{b}_{s,i}$ can be used to ease the calculation of $\hat{\mathbf{c}}_j$, but more importantly; $\mathbf{s}_j$ and $\mathbf{t}_j$ would lie in (or at least be very close to) the space spanned by $\mathbf{\Phi}_{c,j}$. This means that $\mathbf{t}_j$ can be used directly to calculate the first update of $\hat{\mathbf{c}}_j$.

This above scheme represents a sequential relaxation of the constraints imposed by the texture model. At the top of the pyramid, i.e. the smallest image, only a weighted sum of the pixels in the original image is modelled, e.g. the mean of 16 pixels if a box filter and a three level pyramid is used. This is thus a more constrained model compared to the largest model, potentially capable of modelling 16 different pixel values.

We propose to generalise this coarse-to-fine resolution strategy by using the propagation scheme above to traverse sequences of AAMs designed to increase model flexibility. Means for doing so include i) gradual inclusion of more deformation modes in the shape, texture and combined models, ii) independent shape and texture models, iii) de-coupling, or exclusion of shape areas, e.g. omitting the convex hull, modelling only the actual shape area.

9.4.5 Initialisation using Pose Priors

Exploiting the convergence radii of the AAM search carries out model initialisation in this study. Within each radius the AAM will converge to a plausible fit. Parameters with variation outside these radii over the training set are semi-exhaustively searched, with a grid spacing less than twice the radius, by starting the AAM search at each grid point. Each search result represents a potential initialisation candidate. To add robustness this is implemented in a candidate scheme, where several candidates compete in an evolutionary process to become the initial configuration. See [205] for the details of our initialisation algorithm.

As the deformation, size, and orientation variation in our data set are within their convergence radii, the initialisation only needs to search the translational

parameters (x, y) using the maximum likelihood estimate of the corpus callosum, i.e. the mean shape and appearance, having mean size and orientation. Further, this search is restricted by the bounding rectangle of all training shapes.

9.5 Experimental Results

To assess the two landmarking methods, the first three principal modes of shape variation are shown in Figure 9.2. Their shares of the total shape variance were 33%, 30% and 29% (manual) and 35%, 25% and 15% (MDL). As these are ordered by decreasing variance it might surprise that the first manual mode seems to deliver less shape change than the second. The cause is clearly illustrated in Figure 9.3, which shows that variation in the manual case almost solely stems from landmark movements *along* the contour. However, in the MDL case deformations are, by design, much closer to being orthogonal to the contour. Figure 9.2 clearly illustrates that shape modes are *not* canonical, unless a very strict and well-defined landmarking scheme is used. The MDL shape model used only 27% of variance required for the manual shape model to represent the same contour variation. Thus, only MDL-based shapes are considered in the following.

Having established a specific and compact shape deformation basis that ensures correspondence of contour subparts between shapes in a well-defined manner, we

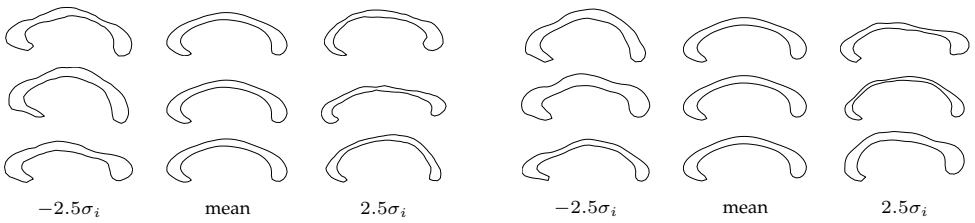


Figure 9.2: The three largest shape deformation modes of $\mathbf{\Phi}_s$ (top-down), manual (left), MDL (right).

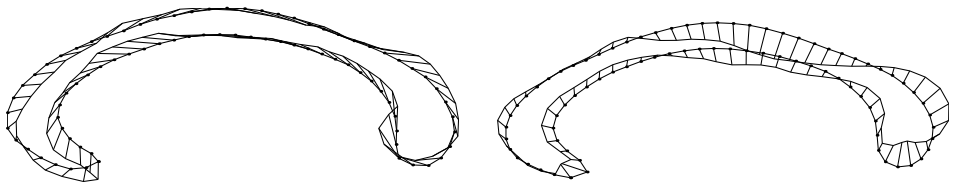


Figure 9.3: Displacement vectors of the largest shape mode in $\mathbf{\Phi}_s$, manual (left), MDL (right).

Table 9.1: Segmentation Results

Case	Whiskers added	Fine-tuning	AAM Seq.	Pt.pt. [pixels] Mean	Pt.pt. [pixels] Median	Pt.crv. [pixels] Mean	Pt.crv. [pixels] Median	Shape overlap Mean	Shape overlap Median
A	No	No	I	2.08±0.32	1.95	0.87±0.07	0.80	0.69±0.02	0.68
B	Yes	No	I	1.24±0.11	1.09	0.73±0.06	0.71	0.76±0.01	0.75
C	Yes	Yes	I	1.14±0.08	1.14	0.68±0.04	0.68	0.76±0.01	0.75
D	Yes	Yes	II	1.13±0.08	1.14	0.67±0.04	0.68	0.76±0.01	0.76
E	Yes	Yes	III	1.07±0.06	1.08	0.60±0.03	0.62	0.78±0.01	0.78
F	Yes	Yes	IV	1.09±0.06	1.11	0.60±0.03	0.60	0.78±0.01	0.79

Sequence I: Size 1:1x1 (one model pixel corresponds to one image pixel);convex hull
Sequence II: Size 1:4x4;convex hull, Size 1:2x2;convex hull, Size 1:1x1;convex hull
Sequence III: (as seq. II), Size 1:1x1;no convex hull
Sequence IV: (as seq. III), Size 1:1x1;no convex hull;no combined model

tested the ability to actively recover these correspondences in unknown images by running an AAM search using the described initialisation method. Each test was carried out using leave-one-out cross-validation. Performance measures were the average distance between i) model and ground truth landmarks (pt.pt.), and ii) model landmark and ground truth contour (pt.crv.). Further, the shape overlap used in [243] has been calculated as the ratio between the ground truth area and the union of the ground truth area and the segmentation result area. The semi-pt.crv. measure used in [43] is the distance between each model landmark and the nearest ground truth landmark. Thus having, pt.crv. and pt.pt. as lower and upper bound, respectively. Texture samples from whiskers influenced the texture model by two thirds. The whisker length was three times the distance between landmark 1 and 2 on the mean shape. Shape, texture and combined eigenspaces were truncated to span 95% of the respective variation.

To improve the AAM search result, the general-purpose optimiser BFGS (see e.g. [96]) was used in case C–F. The cost function was identical to the one used by the AAM search, namely the least squares fit of the model and image texture. AAM sequence designs and results are given in Table 9.1 and shown as box plots in Figure 9.4. Std. errors are symbolised using $\pm$. The pose variations over the data set from the average were 86/130% in size and -13/19 degrees in orientation (min/max). A qualitative impression of the segmentation accuracy is given in Figure 9.5. Images are cropped for display, and do not show the full area searched by the initialisation algorithm.

Fitting the shape model directly to the ground truth shape points gave a lower bound on the pt.pt. accuracy in case F: 0.37/0.84/0.63 pixels (min/max/median). Another lower bound is estimated by a leave-all-in evaluation. On average in case F, this yielded 0.28 pixels pt.pt., 0.18 pixels pt.crv., and a shape overlap of 0.93. In an optimal optimisation process using shape and texture models spanning 100% variation, these would be zero, zero and one.

All results have been generated using an extended version of our public domain C++ AAM implementation [213]. The computation time incl. initialisation in case F was 1.6 seconds on average for each image using a 2.4 GHz Pentium 4 PC. This could easily be reduced to 0.5 seconds by optimisation of the extensions.

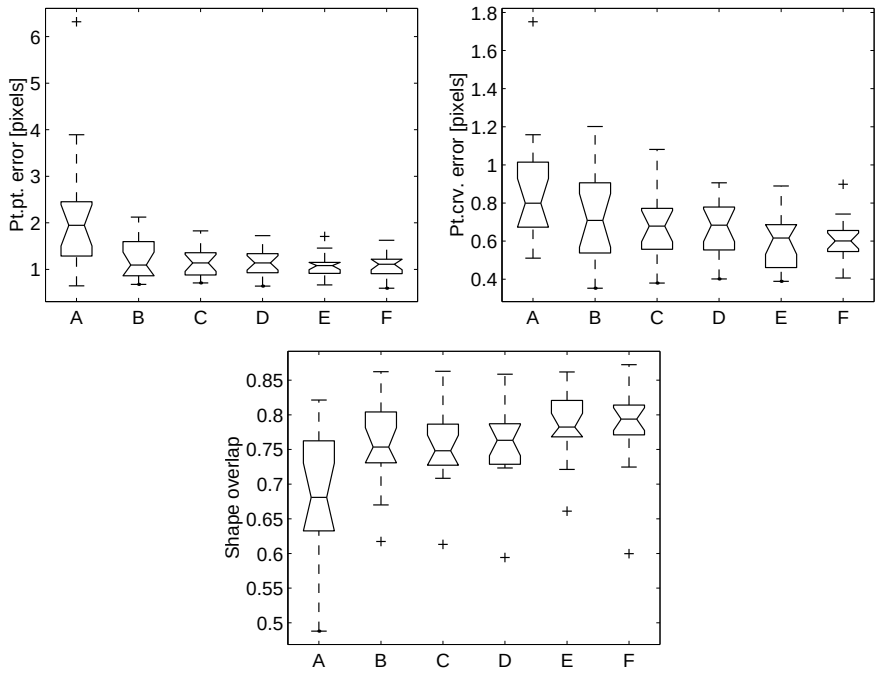


Figure 9.4: Box plots of pt.pt., pt.crv., and shape overlap for each case in Table 9.1. Whiskers are 1.5 IQR at maximum.

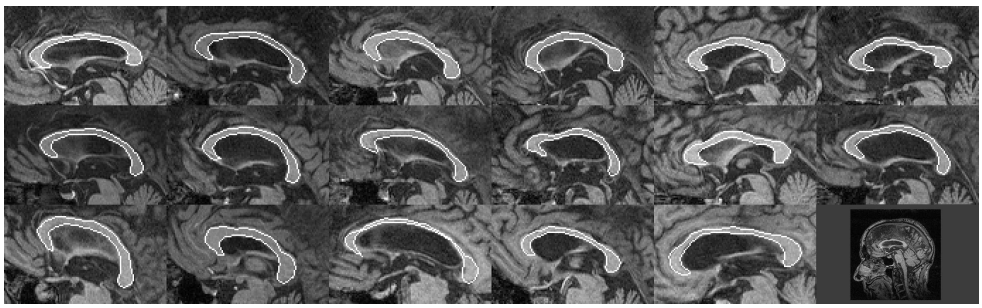


Figure 9.5: Segmentation results, case F (cropped). First image without cropping (lower-right).

9.6 Discussion

From Table 9.1 and Figure 9.4 it is observed that the addition of whiskers markedly improved the segmentation accuracy. The general-purpose optimisation and the AAM sequences brought a modest improvement in accuracy. However, sequence II–IV brought major improvements in initialisation time, due to the sixteen times smaller texture model used for initialisation. Qualitatively, Figure 9.5 shows a very accurate match of each model to the corpus callosum. Quantitatively, these matches are on average only half a pixel from the lower bound. Observation 15 is slightly inaccurate, giving rise to outliers and whiskers in Figure 9.4. The limitations of the shape model are noticeable as subtle discrepancies in observation 3, 11 and 17, all showing small bumps not reachable by the shape model due to the fairly limited number of training shapes. However, considering the large variations present in the data set, we consider the leave-one-out shape model performance impressive.

Alas, due to differences in data sets, comparisons to the related work are not possible. However, this study indicates that our approach easily compares with respect to segmentation accuracy, judging from training set sizes and performance measures.

This study is the only one known to the authors, which establishes compact, well-defined and unique corpus callosum contour correspondences. Shape models based on MDL optimality ensure this. Per se, this uniqueness is a prerequisite for, e.g. proper calculation of splenium areas, and meaningful morphometric regression analysis upon the recovered shape parameters, $\mathbf{b}_s$ in unseen images. Further, only one MR acquisition is required in the presented method, contrary to the work presented in [152]. Finally, this is also the first study that explores the (obvious) combination of MDL shapes models and Active Appearance Models.

9.7 Conclusion

We have presented a both generic and holistic framework, which optimises well-defined measures based on well-described methods, thus avoiding a composition of various techniques, e.g. tissue classification, thresholding, morphologic operators, et cetera. The problem of model initialisation is addressed rendering the presented approach fully automated. Using our freely available AAM implementation, the segmentation time is vastly superior to the reviewed methods, allowing for interactive use, fast analysis of retrospective studies, et cetera.

We conclude that AAMs with the proposed extensions, and in conjunction with MDL-based landmarking, have a great potential as a clinical tool for rapid, accurate and fully automated analysis and segmentation of the corpus callosum in MRI.

Acknowledgements

M.D., M.Sc. Egill Rostrup at Danish Research Centre for Magnetic Resonance initiated this project. The MDL-framework used in this work was created by Davies and co-workers at Division of Imaging Science and Biomedical Engineering – ISBE, Manchester University. We thank Edward Ho at ISBE for his help with the MDL-annotations in an initial stage of this work. The work of M. B. Stegmann was supported by the Danish Medical Research Council, grant no. 52-00-0767.

CHAPTER 10

Segmentation of Anatomical Structures in Chest Radiographs

Bram van Ginneken, Mikkel B. Stegmann, and Marco Loog

Abstract

The task of segmenting the lung fields, the heart, and the clavicles in standard posterior-anterior chest radiographs is considered. Three supervised segmentation methods are compared: active shape models, active appearance models, both first proposed by Cootes et al. [65] and a multi-resolution pixel classification method that employs a multi-scale filter bank of Gaussian derivatives and a k-nearest neighbour classifier. The methods have been tested on a publicly available database of 247 chest radiographs, in which two human observers have manually segmented all objects.

A parameter optimisation for active shape models is presented, and it is shown that this optimisation improves performance significantly. It is demonstrated that the standard active appearance model scheme performs poorly, but large improvements can be obtained by including areas outside the objects into the model.

For lung field segmentation, all methods perform well, with pixel classification giving the best results: a paired t-test showed no significant performance difference between pixel classification and an independent human observer. For heart segmentation, all methods perform comparably, but significantly worse than a human observer. Clavicle segmentation is a hard problem for all methods; best results are obtained with active shape models, but human performance is substantially better.

As an application, the cardiothoracic ratio is computed automatically from the segmentation results. Bland and Altman plots indicate that all methods perform well when compared to the gold standard, with confidence intervals from pixel classification and active appearance modelling very close to those of a human observer.

All results, including the manual segmentations, have been made publicly available to facilitate future comparative studies.

10.1 Introduction

A large amount of literature in the medical image analysis research community is devoted to the topic of segmentation. Many methods have been developed and tested on a wide range of applications. Despite these efforts, or perhaps because of the large number of algorithms that have been proposed, it remains very difficult for a system designer to decide which approach is best suited for a particular segmentation task. Fortunately, there is a growing awareness in the medical image research community that evaluation and performance characterization of segmentation methods is a critical issue [39, 132]. Such evaluations are greatly facilitated by the availability of public image databases with manual annotations on which researchers can test and compare different algorithms. For this study, we have annotated a public database, and have made the manual segmentations available [74].

We compare three methods for segmenting five important anatomical structures in the single most acquired medical image: the standard posterior-anterior (PA) chest radiograph. To this end, these structures – the lung fields, the heart, and the clavicles – have been segmented manually by two observers independently in 247 radiographs from the publicly available JSRT (Japanese Society of Thoracic Radiology) database [194]. The fact that each object has been manually segmented twice allows using one set of manual segmentations as the gold standard and comparing the performance of automatic methods with that of an independent human observer. The web site of the annotated JSRT database [74] allows other researchers to upload the results of other segmentation algorithms applied to the database and we invite the medical image analysis research community to do so.

Accurate segmentation of anatomical structures in chest radiographs is essential for many analysis tasks considered in computer-aided diagnosis. These include various size measurements, the determination of the presence of pulmonary nodules or signs of interstitial lung disease. Knowledge about the location of the clavicles can be used to reduce false positive findings or to detect lesions hiding 'behind a clavicle' more reliably.

The methods considered here are active shape models (ASM) [65, 68], active appearance models (AAM) [54] and pixel classification (PC). ASM is a popular segmentation method, with many internal parameters. We consider how to tune these parameters. AAM has recently found widespread application in medical image segmentation. In this work we use an implementation available in the public domain [213] and compare the standard AAM scheme with an extension in which the surroundings of objects are modelled as well. PC is a classical segmentation method, but the basic concept is so general that it can be implemented in many different ways. We propose an implementation in which both position and local image derivatives are used as input features and show how a multi-resolution implementation and an approximate k-nearest neighbour classifier lead to a relatively fast scheme that yields accurate segmentations.

Each of the methods examined here is *supervised*. This means that example images with the desired output need to be supplied for training. This makes the methods versatile; by supplying different training images and annotations, each method can be applied to many different segmentation tasks, including the ones investigated here. This is in contrast to rule-based schemes that are specifically designed to handle one segmentation task. The usage of these three methods is – contrary to the training process – carried out *without* supervision.

The article is organized as follows. Section 10.2 briefly reviews previous work on segmentation of lung fields, heart and clavicles in chest radiographs. Section 10.3 describes the data. The segmentation methods are presented in Section 10.4. Section 10.5 presents the results, followed by a discussion in Section 10.6. Section 10.7 concludes.

10.2 Previous Work

Segmentation of lung fields in PA chest radiographs has received considerable attention in the literature. Rule-based schemes have been proposed by Li et al. [150], Armato et al. [6], Xu et al. [252, 253], Duryea and Boone [85], Pietka [179], and Brown et al. [44]. Lung segmentation by pixel classification using neural networks has been investigated by McNitt-Gray et al. [157], and Tsujii et al. [236]. Vittitoe et al. [247] developed a pixel classifier for the identification of lung regions using Markov random field modelling. An iterative pixel-based classification method related to Markov random fields was presented in [151]. Van Ginneken and Ter Haar Romeny proposed a hybrid method that combines a rule-based scheme with a pixel classifier [246]. ASM has been used for lung field segmentation in [243, 244].

Segmentation of the outline of the heart has been studied by several researchers, usually with the aim of detecting cardiomegaly (enlarged heart size). For this purpose, only parts of the heart border need to be known. Published methods typically use rule-based schemes, using edge detection and a geometrical model of the heart shape [110, 136, 171, 172, 193].

The segmentation of clavicles in chest radiographs has, to the best of our knowledge, not been studied before.

10.3 Materials

10.3.1 Image Data

The chest radiographs are taken from the JSRT database [194]. This is a publicly available database with 247 PA chest radiographs collected from 13 institutions in Japan and one in the United States. The images were scanned from films to a size

of 2048 $\times$ 2048 pixels, a spatial resolution of .175 mm/pixel and 12 bit grey levels. 154 images contain exactly one pulmonary lung nodule each; the other 93 images contain no lung nodules.

10.3.2 Object Delineation

Each object has been delineated by clicking points along its boundary using a mouse pointer device. These points are connected by straight-line segments. For the ASM and AAM segmentation methods, these contours need to be converted to a fixed number of corresponding points. To this end, several additional, distinguishable points on the contour are clicked by the user, indicating anatomical or other characteristic landmarks. These characteristic points are assumed to correspond. After the complete boundary has been defined, all but the corresponding points are discarded and subsequent points are obtained by equidistantly sampling a certain fixed number of points along the contour between the aforementioned indicated points. This is illustrated in Figure 10.1.

Two observers segmented 5 objects in each image. The first observer was a medical student, the second observer a computer science student. The segmentations of the first observer are taken as gold standard in this study, to which the segmentations of a computer algorithm and the second observer can be compared. The availability of a second observer allows for comparisons between 'human' and 'computer' results, as well as an assessment of the reproducibility of the annotation process.

Observers were allowed to zoom and adjust brightness and image contrast, and could take unlimited time for segmentation. Before the segmentations were made, both observers were instructed by an experienced radiologist. After segmenting all objects, each observer reviewed the results, and adjusted them to correct occasional errors and avoid bias due to learning effects. When in doubt, they reviewed cases with the radiologist. Both observers segmented the images and reviewed the results independently, but they did consult the same radiologist.

10.3.3 Anatomical Structures

In this work we consider the right and left lung, the outline of the heart and the right and left clavicles[1]. It is important to carefully define what is meant by the outline of an anatomical structure in a projection image.

The intensity in each pixel is determined by the attenuation of the radiation by a column of body tissue. One could define the lung fields as the set of pixels for which the radiation has passed through the lung fields. However, this outline is impossible

[1]Note that, by convention, a chest radiograph is displayed as if one is facing the patient. This means that the right lung and clavicle are on the left in the image.

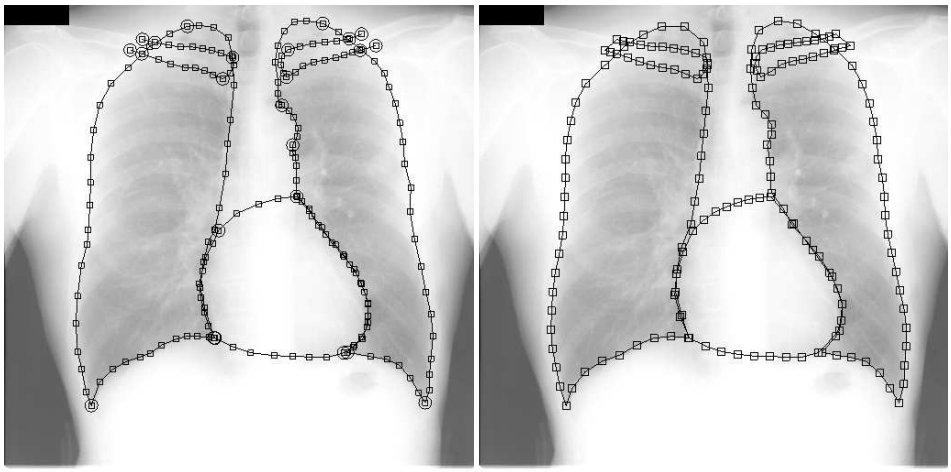


Figure 10.1: Left: the points indicated by the first observer on the first image of the JSRT database to delineate lung fields, the heart, and the clavicles. The anatomical or distinctive points are circled. The right lung contains 3 of these points, the left lung 5, the heart 4 and each clavicle 6. Right: the landmarks interpolated between the anatomical landmarks along the contours indicated on the left for use in the ASM and AAM segmentation method. The total number of landmarks is 166, with 44, 50, 26, 23 and 23 points in right lung, left lung, heart, right clavicle, and left clavicle, respectively.

to determine from a frontal chest radiograph. Therefore we adopt the following definition for the lung fields: any pixel for which radiation passed through the lung, but not through the mediastinum, the heart, structures below the diaphragm, and the aorta. The vena cava superior, when visible, is not considered to be part of the mediastinum.

The heart is defined as those pixels for which radiation passes through the heart. From anatomical knowledge the heart border at the central top and bottom part can be drawn. The great hilar vessels can be assumed to lie on top of the heart.

For the clavicles, only those parts superimposed on the lungs and the rib cage have been indicated. The reason for this is that the peripheral parts of the clavicles are not always visible on a chest radiograph.

Figure 10.1 shows one image and the annotated objects.

10.4 Methods

10.4.1 Active Shape Model Segmentation

The following is a brief description of the ASM segmentation algorithm. The purpose is mainly to point out the free parameters in the scheme; the specific values

Table 10.1: Parameters for ASM. The standard settings are the default values suggested in [65]. The test range was used in pilot experiments to determine optimally tuned settings. The tuned settings are in the last column.

		Standard	Tested	Tuned
Shape model				
align	use Procrustes shape alignment	`true`	`false` / `true`	`false`
f_v	variance to be explained by the shape model	.98	.95 – .995	.995
m	bounds on eigenvalues	3.0	2.0 – 3.0	2.5
Appearance model				
k	points in profile on either side of the point	3	1 – 9	5
L_{max}	resolution levels	4	1 – 6	5
Search algorithm				
n_s	positions to evaluate on either side of point	2	1 – 9	2
N_{max}	max. iterations per level	5	5 – 20	20
p_{close}	convergence criterion	.9	.9 – 1.1*	1.1*

* $p_{close} > 1$ forces N_{max} interations at each level.

for these parameters are listed in Table 10.1. Cootes et al. first introduced the term active shape model in [55] and [68]. However, [68] does not include the gray level appearance model and [55, 68] do not include the multi-resolution ASM scheme. Both of these components are essential to obtain good segmentation results with ASM in practice. Our implementation follows the description of the ASM method given in [65] to which the reader is referred for details.

The ASM scheme consists of three elements: a global shape model, a local, multi-resolution appearance model, and a multi-resolution search algorithm.

A set of objects in a training image is described by n corresponding points. These points are stored in a shape vector, $\mathbf{x} = [\, x_1\ y_1\ \ldots\ x_n\ y_n\,]^\mathsf{T}$. A set of these vectors can be aligned by translating, rotating and scaling them so as to minimize the sum of squared distances between the points (generalised Procrustes alignment [65, 107]). Alignment can also be omitted, which will include the variation in size and pose into the point distribution model, which is subsequently constructed. Let $\overline{\mathbf{x}}$ denote the mean shape. The t principal components (modes of variation in the shape model) of the covariance matrix of the shape vectors are computed. The value of t is determined by f_v, the amount of variation in the training shapes one wants to explain. Shapes can now be written as

$$\mathbf{x} = \overline{\mathbf{x}} + \mathbf{\Phi}_x \mathbf{b}_x, \tag{10.1}$$

where $\mathbf{\Phi}_x$ contains the modes of variation of the shape model and $\mathbf{b}_x$ holds the shape parameters. During the ASM search it is required to fit the shape model to a set of landmarks. This is done by projecting the set of landmarks on the eigenvectors in $\mathbf{\Phi}_x$ and truncating each projection in $\mathbf{b}_x$ to m times the standard deviation in that direction.

A local appearance model is constructed for each landmark. On either side of the contour at which the landmark is located, k pixels are sampled using a fixed step size of 1 pixel, which gives profiles of length $2k + 1$. Cootes et al. propose to use the normalized first derivatives of these profiles [65]. The derivatives are computed using finite differences; the normalization is such that the sum of absolute values equals 1. Note that this requires a notion of connectivity between the landmark points from which the direction perpendicular to the contour can be computed.

As a measure for the goodness of fit of a pixel profile encountered during search, the Mahalanobis distance to the set of profiles sampled from the training set is computed. These profile models are constructed for L_{max} resolutions. A standard image pyramid [45] is used.

The search algorithm is a simple iterative scheme initialised by the mean shape. Each landmark is moved along the direction perpendicular to the contour to n_s positions on either side, evaluating a total of $2n_s + 1$ positions. The landmark is put at the position with the lowest Mahalanobis distance. After moving all landmarks, the shape model is fitted to the displaced points, yielding an updated segmentation. When a proportion p_{close} of points ends up within $n_s/2$ of its previous position, or when N_{max} iterations have been made, the search moves to the next resolution level, or ends. The highest resolution level in our experiments was 256 $\times$ 256 pixels. The use of higher resolutions did not improve performance.

In [65] suitable values are suggested for all parameters in the ASM scheme. They are listed in Table 10.1 and they have been used in the experiments, referred to as 'ASM default'. In order to investigate the effect of different settings, we performed pilot experiments on a small test set (to keep computation time within reasonable bounds) and varied all settings within a sensible range, also given in Table 10.1. The overall best setting was kept (last column in Table 10.1) and also used in the experiments, referred to as 'ASM tuned'.

10.4.2 Active Appearance Models

The active appearance model (AAM) segmentation and image interpretation method [54] has recently received a considerable amount of attention in the image analysis community [213]. AAM uses the same input as ASM, a set of training images in which a set of corresponding points has been indicated.

The major difference to ASM is that an AAM considers all objects pixels, compared to the border representation from ASM, in a combined model of shape and appearance. The search algorithm is also different. This section will summarize the

traditional AAM framework, list the parameter settings and describe alterations we applied for this segmentation task. Our implementation was based on the freely available C++ AAM implementation described in [213].

An AAM is a generative model, which is capable of synthesising images of a given object class. By estimating a compact and specific basis from a training set, model parameters can be adjusted to fit unseen images and hence perform both image interpretation and segmentation. The modelled object properties are shape – using the shape vectors $\mathbf{x}$ – and pixel intensities (called *texture*), denoted by $\mathbf{t}$. As in ASM, variability is modelled by means of principal component analyses (PCA). Prior to PCA modelling, shapes are Procrustes aligned and textures are warped into a shape-free reference frame and sampled. Usually only the convex hull of the shape is included into the texture model. It is also possible to model the inside of every closed contour. New instances for the shape can be generated by Equation 10.1, and, similarly, we have for the texture

$$\mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_t \mathbf{b}_t \tag{10.2}$$

where $\bar{\mathbf{t}}$ denotes the mean texture, $\mathbf{\Phi}_t$ are eigenvectors of the texture dispersions (both estimated from the training set) and $\mathbf{b}_t$ holds the texture model parameters. To recover any correlation between shape and texture and obtain a combined parameterisation, $\mathbf{c}$, the values of $\mathbf{b}_x$ and $\mathbf{b}_t$ are combined in a third PCA,

$$\begin{bmatrix} \mathbf{W}_x \mathbf{\Phi}_x^{\mathrm{T}} (\mathbf{x} - \bar{\mathbf{x}}) \\ \mathbf{\Phi}_t^{\mathrm{T}} (\mathbf{t} - \bar{\mathbf{t}}) \end{bmatrix} = \begin{bmatrix} \mathbf{W}_x \mathbf{b}_x \\ \mathbf{b}_t \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_{c,x} \\ \mathbf{\Phi}_{c,t} \end{bmatrix} \mathbf{c} = \mathbf{\Phi}_c \mathbf{c}. \tag{10.3}$$

Here, $\mathbf{W}_x$ is a diagonal matrix weighting pixel distances against intensities.

Synthetic examples, parameterized by $\mathbf{c}$, are generated by

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{\Phi}_x \mathbf{W}_x^{-1} \mathbf{\Phi}_{c,x} \mathbf{c}$$

and

$$\mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_t \mathbf{\Phi}_{c,t} \mathbf{c}$$

and rendered into an image by warping the pixel intensities of $\mathbf{t}$ into the geometry of the shape $\mathbf{x}$.

Using an iterative updating scheme the model parameters in $\mathbf{c}$ can be fitted rapidly to unseen images using the L_2-norm as a cost function. See [54, 65] for further details. As in ASM, a multi-resolution pyramid is used.

Parameter Settings

Segmentation experiments were carried out in a two-level image pyramid (128 $\times$ 128 and 256 $\times$ 256 pixels). The use of higher resolution did not improve performance.

The model was automatically initialised on the top level by a sparse sampling in the observed distribution of training set pose. This sparseness is obtained by considering the convergence radius of each model parameter (inspired by [65]), thus avoiding any unnecessary sampling. Since rotation variation was completely covered by the convergence radius, no sampling was performed in this parameter. From the training data, it was estimated that the model should converge if initialised in a 2 by 2 grid around the mean position. Further, due to the variation in size over the training set, each of these four searches was started at 90%, 100%, and 110% of the mean size, respectively. Thus, 12 AAM searches were executed in each image and the search producing the best model-to-image fit was selected.

Both shape, texture and combined models were truncated at $f_v = .98$, thus including 98% of the variance. Bounds m on the combined eigenvalues were three standard deviations. Model searches had a limit of 30 iterations at each pyramid level. AAM parameter update matrices for pose and model parameter were calculated using Jacobian matrices. These were estimated using every 15th training case. Parameter displacements were as follows: *model parameters*: $\pm 0.5\sigma_i$, $\pm 0.25\sigma_i$ (σ_i denotes the standard deviation of i^{th} parameter), *x-y position*: $\pm 2\%$, $\pm 5\%$ (of the width and height), *scale*: $\pm 2\%$, $\pm 5\%$, *rotation*: $\pm 2\%$, $\pm 5\%$ degrees. Displacements were carried on sequentially; i.e. one experiment for each displacement setting. The details of this process can be found in [54], and are further expanded in [213].

AAM with Whiskers

In this particular application of segmenting chest radiographs, the objects in question are best characterised by their borders. They do not have much distinct and consistent interior features; the lungs show a pattern of ribs and vasculature but the location of these structures relative to the points that make up the shape is not fixed, the heart is dense, but opaque, and no distinct structures can be observed. This behaviour is common to many medical image analysis problems and poses a problem to the original AAM formulation where only the object's interior is included into the texture model. This means that the cost function can have a minimum when the model is completely inside the actual object. To avoid this, information about the contour edges needs to be included into the texture model. We use the straightforward approach from ASMs; namely to add contour normals pointing outwards on each object. These normals are in this context denoted *whiskers* and are added implicitly during texture sampling with a scale relative to the current shape size. Texture samples obtained by sampling along whiskers are now concatenated to the texture vector, $\mathbf{t}$, with a uniform weight relating these to the conventional AAM tex-

ture samples obtained inside every closed contour. This provides a simple weighted method for modelling object proximity in an AAM. Unfortunately, this also introduces two additional free parameters, the length of the whiskers and the weighting of whisker samples.

The following parameters were chosen: whisker length was equal to distance between landmark 1 and 2 on the mean shape (sized to mean size) and texture samples from whiskers influenced the texture model with the same weight as the normal, interior texture samples. Pilot studies showed that moderate changes from the parameter set chosen above had no significant impact on the accuracy.

Refinement of AAM Search Results

AAMs provide a very fast search regime for matching the model to an unseen image using prior knowledge derived from the training set. However, due to the approximate nature this process will not always converge to the minimum of the cost function. A pragmatic solution to this problem is to refine the model fit by using a general-purpose optimisation method. Assuming that the AAM search brings the model close to the actual minimum, this approach is feasible wrt. computation, despite the typical high-dimensional parameter space of AAMs. We have used a gradient-based method for this application. The cost function remained unchanged; the L_2-norm between model and image texture. It was optimised by a quasi-Newton method using the BFGS (Broyden, Fletcher, Goldfarb and Shanno) update of the Hessian, see e.g. [96]. Alternatively, to avoid spurious minima, a random-sampling method such as Simulated Annealing can be employed. Refinement of AAMs has previously been employed to improve the model fit in cardiac and brain MRI by Stegmann et al. [214].

10.4.3 Pixel Classification

Pixel classification (PC) is an established technique for image segmentation. It enjoys popularity in many areas of computer vision, e.g. remote sensing [183]. Within medical imaging, it has been used extensively in multi spectral MR segmentation [22]. A recent example of an application to 3D MR brain segmentation can be found in [47]. In chest radiograph segmentation it has been used before in [157, 246] and, in the context of Markov random field segmentation in [247].

This section describes a general multi-resolution implementation of PC that we developed along with the components and parameters used for this particular segmentation problem.

General Algorithm

PC is composed of a training and a test stage. The training stage consists of

1. Choose a working resolution. Obtain a copy of each training image at this working resolution.
2. Choose a number of samples (positions) in each training image.
3. Compute a set of features (the input) for each sample. Possible features are the grey level value at that position or in the surroundings, filter outputs, and position values. Associate an output with each sample. This output lists to which classes this position belongs. Note that in our case pixels can belong to multiple classes simultaneously (e.g. left clavicle and left lung field).
4. (Optional) Compute a suitable transformation for the feature vectors. Examples of transformations are normalization, feature selection, feature extraction by PCA or whitening, or non-linear transformation to create new, extra features.
5. Train a classifier with the input feature vectors and the output; this classifier can map new input to output. In this work we require that the classifier can compute the posterior probability (the probability, given the input features) that a pixel belongs to each object class.

The test stage consists of

1. Obtain a copy of the test image at the working resolution.
2. Compute the features for each pixel in the image.
3. (Optional) Apply the transformation to each feature vector.
4. Obtain the posterior probabilities that the pixel belongs to each class, using the transformed feature vectors and the trained classifier.
5. A binary segmentation for each object is obtained by thresholding the output at .5. Optionally, post-processing operations can be applied before and after binarisation.

Multi-resolution PC

In the multi-resolution PC method, training is performed for a range of working resolutions. The test stage begins at the coarsest resolution and stores the posterior probabilities p_i for each class i and also stores $p_{min} = \min_i(p_i^*)$ where $p_i^* = \max(p_i, 1 - p_i)$, the chance that the pixel belongs to class i, or not, whichever is more likely. If p_{min} is close to 1, the classifier is confident about all the class labels of that pixel. If it is close to .5, the classifier is unsure about at least one of the labels. The test stage continues at the next resolution level. In this level, the number of pixels is larger. The p_{min} values are linearly interpolated from the previous level, and

only if $p_{min} < T$, a pixel is reclassified. Otherwise, the interpolated posterior labels of the coarser level are taken. This process continues until the finest resolution has been processed.

The rationale behind this strategy is that classification (step 4 in the test algorithm given above) is usually the computationally most expensive operation. As low resolution images contain less pixels, and in many applications a large area of the image is often easy to classify (p_{min} close to 1), using the multi-resolution scheme can speed up the PC process considerably. Estimating *all* p_i at a next level if *any* p_i^* is below T may seem superfluous. For the classifier of our choice, however, this is as expensive as only estimating those p_i for which $p_i^* < T$.

Lower resolution images were created with the Gaussian pyramid [45], as is done in ASM. In the experiments just two resolution levels were considered, where the images were reduced to 128 by 128 and 256 by 256 pixels. The threshold T was conservatively set to .99. Higher resolutions increased computation time but did not improve performance; more lower resolution levels slightly decreased performance at hardly any computational gain.

Samples and Features

A rectangular grid of 64 by 64 pixels was placed over each training image to extract 4096 samples per image.

Spatial features, the (x, y) coordinates in the image, are used, because the structures we aim to segment have a characteristic location within a chest radiograph. Additionally, the output of Gaussian derivative filters of up to second order ($L, L_x, L_y, L_{xx}, L_{yy}, L_{xy}$) at five scales ($\sigma = 1, 2, 4, 8, 16$ pixels at the current resolution) are used to characterize local image structure. Finally, the gray value in the original images was taken as a feature.

This set of features was computed at each resolution level. As the pixel size of images is different at each level, the scale of the filters is different as well, to the effect that large apertures are used for a first coarse segmentation and finer apertures subsequently segment the finer details at higher resolution levels.

Classifier and Feature Transformations

The effect of feature transformations is closely related to the choice of classifier. A k-nearest neighbour (kNN) classifier was used, with $k = 15$. The kNN classifier has the attractive property that, under certain statistical assumptions and in the case of infinite training data, the conditional error is $(1+1/k)R^*$, where R^* is the minimally achievable Bayes error [83]. Mount and Arya's tree-based kNN implementation [8], was employed which allows for a considerable speed up of the classification by calculating an approximate solution. The approximation is controlled by a variable ϵ, which means that the approximate nearest neighbours which the algorithm finds,

are no more than $(1 + \epsilon)$ the distance away from the query point than the actual nearest neighbours are [8]). ϵ was set to 2. This did not lead to a decrease in accuracy as compared to exact kNN with $\epsilon = 0$.

In pilot experiments, various feature selection and feature extraction methods were tested, but they did not yield a significant performance increase. Eventually, we only applied normalization, which means that a scaling factor per feature is determined so that each feature has unit variance in the training set.

An additional advantage of the kNN classifier, already hinted at above, is that the posterior probability for each object class can be determined using only one neighbour search. Note that the combination of $k = 15$ and $T = .99$ means that pixels are only not reclassified at a finer resolution level if all k neighbours have the same class label configuration (as $14/15 < .99$).

Post-processing

The obvious way to turn the soft classification p_i into binary masks is thresholding at a posterior probability of .5. However, this does not ensure connected objects; segmentations will often contain clouds of isolated pixels near the object's boundary. To ensure that the segmentation for each structure yields a single connected object, a simple post-processing procedure was developed. This procedure is the same for each object considered.

First the soft output is blurred with σ = 0.7 mm. This reduces the grainy appearance at object boundaries and can be interpreted as pooling of local evidence. Subsequently the largest connected object is selected, and holes in this object are filled.

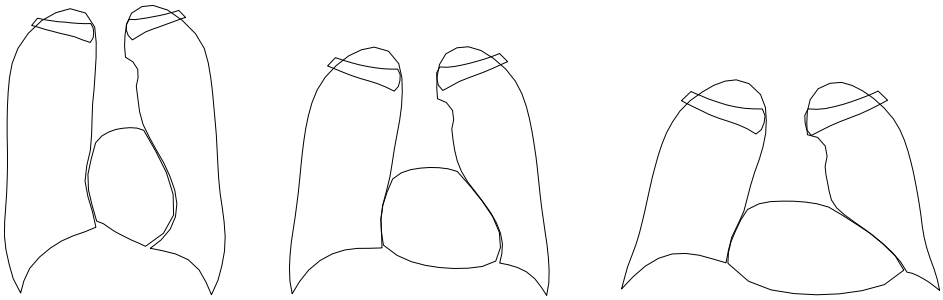
10.5 Experimental Results

10.5.1 Point Distribution Model

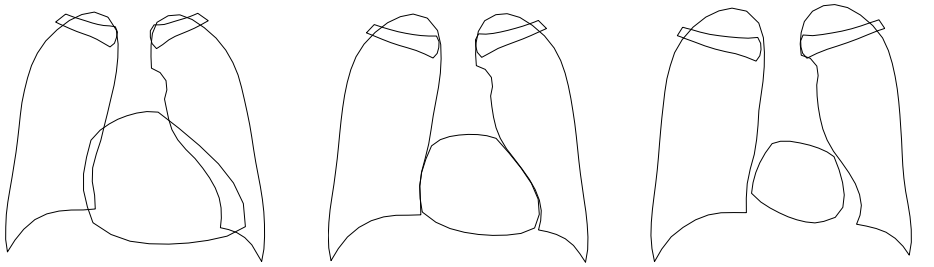
The analysis of the shape vectors, $\{\mathbf{x}\}$, gives insight in the typical variations in shape of lungs, heart and clavicles that occur in chest radiographs, and their correlation. This is an interesting result in its own right, and therefore the first few modes of variation are displayed in Figure 10.2. In Figure 10.3 the spread of each model point after Procrustes alignment is displayed. This is another way of visualizing which parts of the objects exhibit most shape variation.

10.5.2 Folds

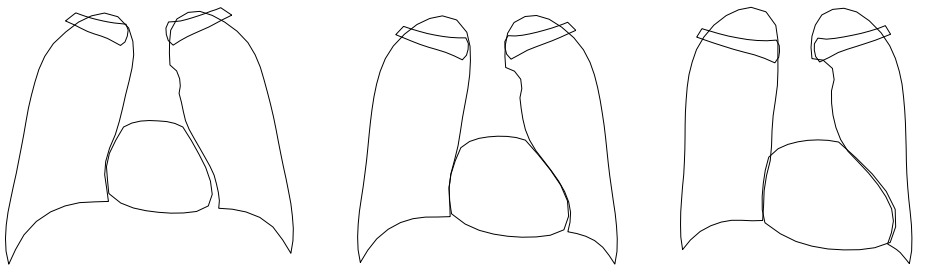
The 247 cases in the JSRT database were split in two folds. One fold contained all 124 odd numbered images in the JSRT database. The other fold contained the 123



(a) $b_1 = -3\sqrt{\lambda_1}$ (b) $b_1 = 0$ (c) $b_1 = +3\sqrt{\lambda_1}$



(d) $b_2 = -3\sqrt{\lambda_2}$ (e) $b_2 = 0$ (f) $b_2 = +3\sqrt{\lambda_2}$



(g) $b_3 = -3\sqrt{\lambda_3}$ (h) $b_3 = 0$ (i) $b_3 = +3\sqrt{\lambda_3}$

Figure 10.2: Mean shape deformation obtained by varying the first three modes of the shape model between -3 and +3 standard deviations (denoted by $\sqrt{\lambda_i}$).

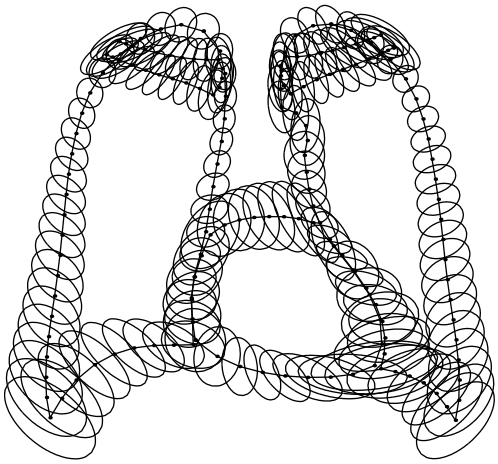


Figure 10.3: Independent principal component analysis for each model point after Procrustes alignment.

even numbered images. This division ensured that both folds contained an equal amount of normal cases and cases with a lung nodule. Images in one fold were segmented with the images in the other fold as training set, and vice versa.

10.5.3 Performance Measure

To measure the performance of a segmentation algorithm, a 'goodness' index is required. For a two-class segmentation problem, one can distinguish true positive (TP) area (correctly classified as object), false positive (FP) area (classified as object, but in fact background), false negative (FN) area (classified as background, but in fact object), and true negative (TN) area (correctly classified as background). From these values, measures such as accuracy, sensitivity, specificity, kappa and overlap can be computed. In this work we use the intersection divided by union as an overlap measure, given by

$$\Omega = \frac{TP}{TP + FP + FN}. \tag{10.4}$$

This is a well accepted measure, but one should be aware that objects that are small or have a complex shape usually achieve a lower Ω than larger objects [104].

For comparisons between methods, paired t-tests were used. Differences are considered significant if $P < 0.05$.

10.5.4 Evaluated Methods

The five objects in each of the 247 images were segmented with 12 methods in total:

- First of all, the segmentations of the second human observer were used to compare computerized methods with human performance.
- As a reference method, we computed the performance when the mean shape of each object is taken as segmentation, independent of the actual image contents. Clearly any method should outperform this 'a priori' segmentation.
- Two ASM systems were employed; ASM with the 'default' settings and the 'tuned' settings given in Table 10.1.
- For AAM, three systems were evaluated: the 'standard' system (Section 10.4.2); the version with whiskers added (Section 10.4.2) and finally, the system with whiskers refined by BFGS (Section 10.4.2).
- The results for pixel classification are given both with and without post-processing (Section 10.4.3).
- To obtain upper bounds for the performance of ASM and AAM systems, the tuned ASM method was run, initialised from the gold standard; the ASM shape model was fitted directly to the gold standard and the AAM method with whiskers was method was run, initialised from the gold standard. Note that these result are supplied only for reference, obviously these system cannot be used in practice as they require the gold standard to be known.

10.5.5 Segmentation Results

Results of all 9 systems are listed in Table 10.2. In Figure 10.4 the quartiles are shown graphically in box plots. For results of the three ASM/AAM systems that were started from the ground truth see Table 10.3. In general, the best results *per system* were obtained by the tuned ASM system, the AAM system with whiskers and BFGS refinement added and the PC system with post processing. However, for clavicle segmentation, post-processing did not significantly improve PC segmentation and use of the AAM BFGS refinement did not improve upon AAM with whiskers only.

For lung field segmentation, PC clearly outperforms ASM and AAM. Interestingly, ASM and AAM perform comparably for the right lung field, but ASM is significantly better than AAM for the left lung field. There is no significant difference between PC and the human observer for both lung fields.

For heart segmentation, performance of the human observer is lowest for all objects, but significantly better than any computer methods. AAM, PC and ASM are all close; the only significant difference is that between AAM and ASM.

Clavicle segmentation proves to be a hard problem for any of the methods. The human observer greatly outperforms any computer method. Best results are obtained with ASM, followed by AAM and PC. Only the difference between the latter two is not significant.

Table 10.2: Segmentation results for the left clavicle, for each system considered. All results are in terms of the overlap Ω, as defined in Equation 10.4.

Right lung	$\mu \pm \sigma$	Min	Q1	Median	Q2	Max
human observer	.951 ± .021	.851	.942	.955	.964	.981
mean shape	.740 ± .087	.494	.683	.742	.805	.907
ASM default	.908 ± .061	.491	.900	.927	.940	.968
ASM tuned	.931 ± .032	.732	.925	.939	.950	.972
AAM standard	.859 ± .096	.000	.835	.880	.917	.963
AAM whiskers	.918 ± .037	.706	.907	.927	.942	.960
AAM whiskers BFGS	.930 ± .030	.680	.923	.936	.948	.970
PC, not post-processed	.944 ± .030	.737	.941	.953	.960	.972
PC, post-processed	.951 ± .021	.816	.946	.956	.963	.973
Left lung	$\mu \pm \sigma$	Min	Q1	Median	Q2	Max
human observer	.941 ± .026	.699	.935	.945	.955	.977
mean shape	.687 ± .093	.348	.624	.689	.750	.908
ASM default	.898 ± .063	.578	.878	.920	.940	.964
ASM tuned	.922 ± .046	.576	.914	.936	.948	.968
AAM standard	.835 ± .106	.034	.797	.870	.899	.950
AAM whiskers	.908 ± .043	.686	.895	.919	.937	.961
AAM whiskers BFGS	.914 ± .039	.684	.904	.926	.938	.966
PC, not post-processed	.932 ± .036	.690	.923	.942	.954	.969
PC, post-processed	.939 ± .032	.687	.933	.946	.958	.975
Heart	$\mu \pm \sigma$	Min	Q1	Median	Q2	Max
human observer	.878 ± .054	.571	.843	.888	.916	.965
mean shape	.643 ± .147	.221	.550	.665	.754	.921
ASM default	.793 ± .119	.220	.755	.824	.872	.954
ASM tuned	.814 ± .076	.520	.770	.827	.873	.938
AAM standard	.775 ± .135	.026	.733	.806	.860	.947
AAM whiskers	.813 ± .080	.489	.770	.823	.874	.938
AAM whiskers BFGS	.834 ± .070	.510	.791	.845	.882	.967
PC, not post-processed	.811 ± .077	.497	.769	.832	.862	.914
PC, post-processed	.824 ± .077	.500	.783	.844	.877	.932
Right clavicle	$\mu \pm \sigma$	Min	Q1	Median	Q2	Max
human observer	.890 ± .045	.702	.868	.900	.923	.962
mean shape	.317 ± .236	.000	.079	.320	.503	.851
ASM default	.693 ± .146	.000	.633	.727	.789	.897
ASM tuned	.732 ± .143	.127	.675	.773	.830	.925
AAM standard	518 ± .239	.000	.393	.582	.696	.849
AAM whiskers	.623 ± .180	.000	.570	.665	.743	.865
AAM whiskers BFGS	.639 ± .177	.000	.566	.687	.766	.899
PC, not post-processed	.619 ± .115	.210	.554	.636	.711	.851
PC, post-processed	.616 ± .140	.000	.549	.648	.718	.865
Left clavicle	$\mu \pm \sigma$	Min	Q1	Median	Q2	Max
human observer	.902 ± .044	.695	.885	.916	.932	.958
mean shape	.289 ± .233	.000	.068	.245	.489	.787
ASM default	.687 ± .164	.000	.639	.733	.795	.896
ASM tuned	.736 ± .144	.059	.706	.781	.832	.932
AAM standard	.492 ± .247	.000	.403	.550	.680	.863
AAM whiskers	.626 ± .185	.000	.576	.667	.741	.889
AAM whiskers BFGS	.644 ± .189	.000	.583	.710	.760	.888
PC, not post-processed	.617 ± .115	.157	.550	.628	.700	.831
PC, post-processed	.614 ± .145	.135	.536	.643	.717	.867

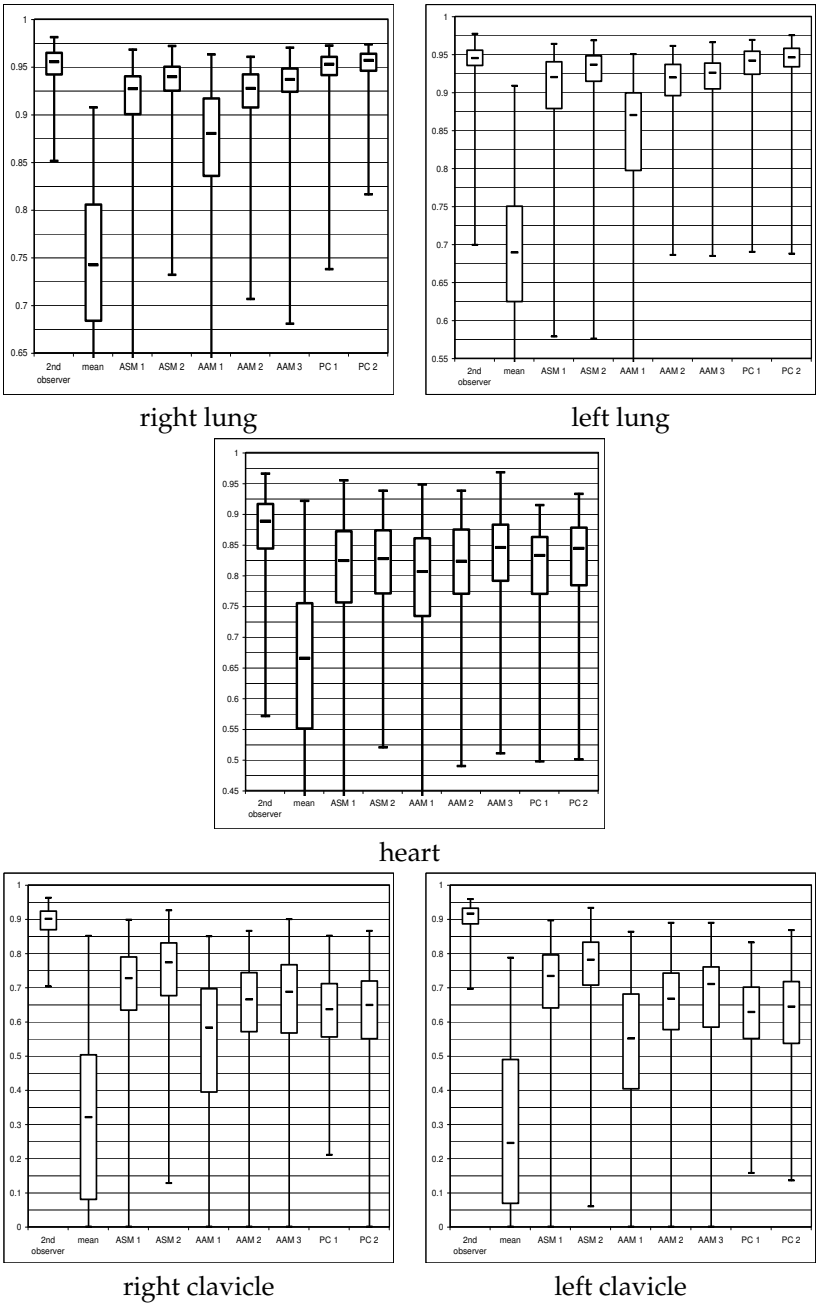


Figure 10.4: Box plots that display the quartiles of all systems for each of the objects segmented. The corresponding numerical values are listed in Table 10.2.

Table 10.3: Segmentation results per object, using the tuned ASM system initialised with the gold standard, fitting the shape model from the ASM system directly to the gold standard, and the AAM whiskers system initialised with the gold standard. These results provide upper bounds for ASM and AAM systems. All results are in terms of the overlap Ω, as defined in Equation 10.4.

Right lung	$\mu \pm \sigma$	Min	Q1	Median	Q2	Max
ASM GS	$.933 \pm .032$	.669	.925	.941	.950	.968
shape model fit	$.949 \pm .019$	.790	.943	.953	.959	.972
AAM GS	$.938 \pm .016$	.878	.930	.941	.951	.971
Left lung	$\mu \pm \sigma$	Min	Q1	Median	Q2	Max
ASM GS	$.926 \pm .042$	.579	.920	.938	.949	.969
shape model fit	$.942 \pm .026$	.726	.934	.949	.957	.971
AAM GS	$.929 \pm .025$	.789	.920	.932	.946	.965
Heart	$\mu \pm \sigma$	Min	Q1	Median	Q2	Max
ASM GS	$.817 \pm .082$	.498	.771	.821	.881	.950
shape model fit	$.943 \pm .039$	.442	.935	.949	.959	.979
AAM GS	$.882 \pm .046$	.656	.858	.887	.917	.962
Right clavicle	$\mu \pm \sigma$	Min	Q1	Median	Q2	Max
ASM GS	$.737 \pm .134$	.184	.696	.766	.827	.914
shape model fit	$.822 \pm .065$	.357	.794	.833	.862	.944
AAM GS	$.721 \pm .090$	.335	.666	.732	.795	.889
Left clavicle	$\mu \pm \sigma$	Min	Q1	Median	Q2	Max
ASM GS	$.738 \pm .137$	.201	.704	.779	.825	.899
shape model fit	$.812 \pm .073$	.309	.793	.825	.853	.909
AAM GS	$.729 \pm .097$	.180	.682	.742	.801	.896

Figure 10.5 shows the results for four cases. These images were selected in the following way. For each image, the overlap of each object when segmented with the ASM, AAM and PC system was averaged. All images were sorted on this 'overall overlap average'. The images ranking #1, #82, #165 and #247 are displayed, corresponding to an easy, relatively easy, relatively hard and a hard case, respectively.

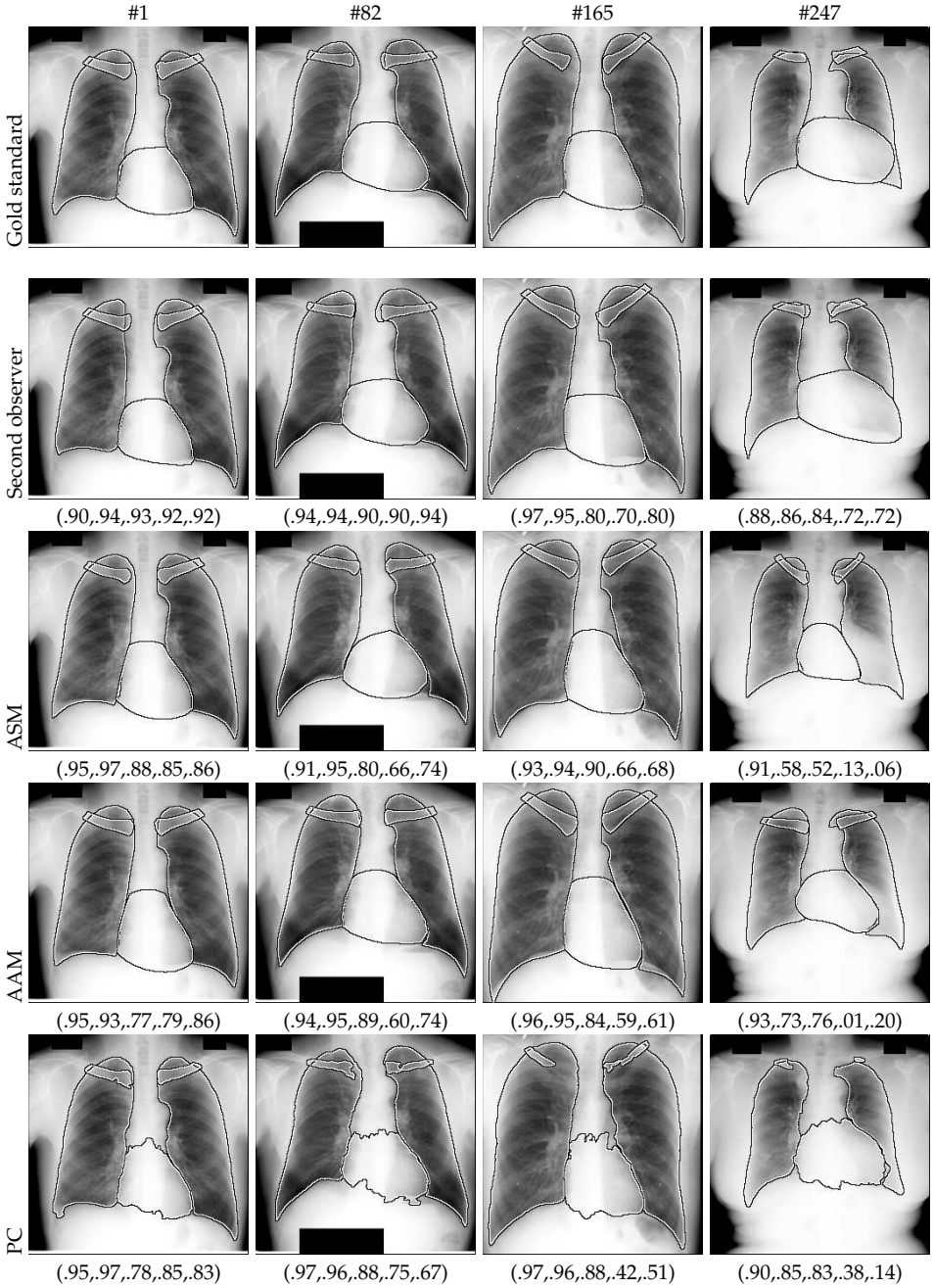


Figure 10.5: Segmentation results for the best ASM, AAM and PC system, compared with manual segmentations, for four cases. Cases were selected according to their ranking (see text for details). Below each image the overlap Ω is listed for the right lung, left lung, heart, right clavicle and left clavicle, respectively.

10.5.6 Computation of the Cardiothoracic Ratio

A segmentation in itself is hardly ever the final outcome of a computer analysis in medical imaging. The ultimate 'goodness' index for segmentation is its usefulness for subsequent processing. One important diagnostic measure that can be directly calculated from a segmentation of lungs and heart in a chest radiograph is the cardiothoracic ratio (CTR), defined as the ratio of the transverse diameter of the heart to the transverse diameter of the thorax. A ratio above 0.5 is generally considered a sign of cardiomegaly, and this test is used frequently in clinical practice and clinical research (e.g. [135]). Automatic computation of the CTR has been investigated before [172, 193].

We computed the CTR from the gold standard, and compared the results with the second observer and the best ASM, AAM and PC systems. Bland and Altman plots [27] are given in Figure 10.6 together with the mean absolute difference and the 95% confidence intervals. Note that the confidence interval is tighter for the PC system than for the second observer. This is due to an outlier, though. If that outlier is

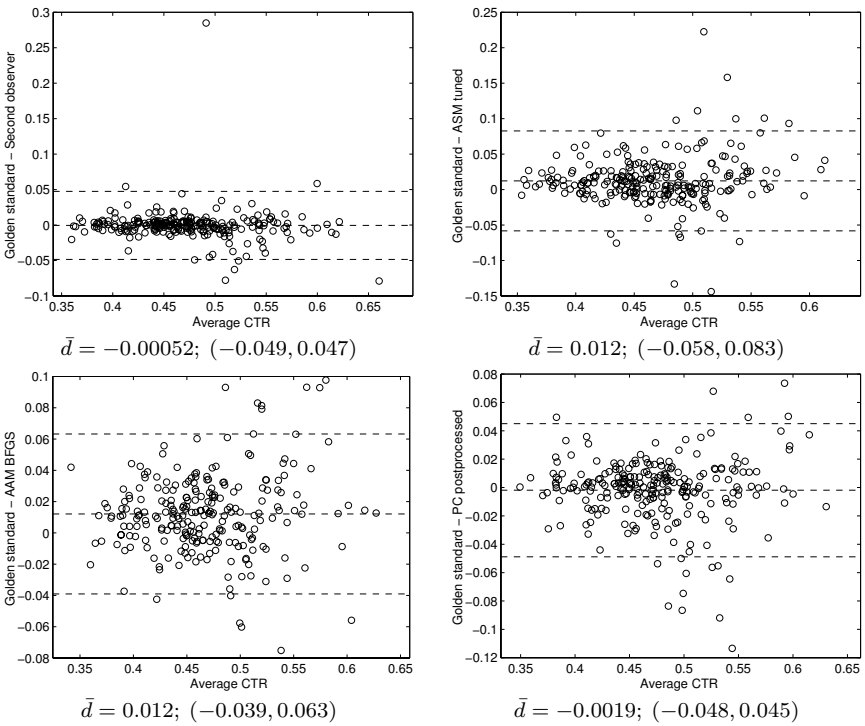


Figure 10.6: Bland and Altman plots of the cardiothoracic ratio computed from the gold standard versus the second observer, ASM, AAM and PC. In the graphs and below them the mean difference and the 95% confidence intervals $(\bar{d} - 2\sigma, \bar{d} + 2\sigma)$ are given.

removed, the confidence interval for the second observer shrinks to $(-0.033, 0.030)$. The confidence interval of PC is tighter than that of AAM, which is tighter than ASM. From the Bland and Altman plots it can be appreciated that there is more often substantial disagreement between the gold standard and computerized measures for cases with a large CTR.

10.5.7 Computation Times

The ASM and PC segmentation were performed on a 2.8 GHz Intel PC with 2 GB RAM. The AAM experiments were carried out on a 1.1Ghz Athlon PC equipped with 768 MB RAM. All implementations were in C++, and in all cases there is room for optimisations. Computation time required for segmenting a single image was around 1 s for ASM, 30 s for PC, and 3 s for AAM.

10.6 Discussion

Some of the presented results obtained by computer algorithms are very close to human performance. Therefore we start this discussion by considering the limitations of manual segmentations, which were used to determine the gold standard, and discuss the representativity of the data. Then the results for lung segmentation, heart segmentation, clavicle segmentation and the automatic determination of the CTR are discussed. After pointing out the fundamental differences between pixel classification, active shape and appearance models, we briefly consider some possibilities for improvements in each of the three methods.

Accuracy of the Gold Standard and Representativeness of the Data

Supervised segmentation methods require training data for which the 'truth' is available, and their performance will therefore depend on the quality of this 'truth'. In this work, manual segmentations from a single observer are taken as gold standard. There are two types of inaccuracies in the gold standard. Occasionally, the observer may misinterpret the superimposed shadows and follow the wrong edge or line in the image. Such *interpretation errors* occur mainly along the mediastinum and heart border and in some cases when edges of clavicles and ribs create a confusing pattern. Interpretation errors can lead to relatively large distances between boundaries drawn by the two observers. The outlier for the second observer versus the gold standard in Figure 10.6 is an example of an interpretation error (of the second observer, as was judged retrospectively). Interpretation errors are more likely to occur when the image contains pathology or unusual anatomy. Some errors of computer algorithms could be considered interpretation errors as well such as allotting areas of the stomach or bowels to the left lung, which happens when there is a lot of air in the stomach and the diaphragm below the left lung has a line-like

instead of an edge-like appearance. Another example is following the wrong edge for the border between heart and left lung, of which some examples can be seen in Figure 10.5.

The second type of inaccuracy could be described as *measurement error*. Clicking points along the boundary of hundreds of objects is a straining task for human operators; inevitably small errors are made. It is possible that supervised computerized methods such as the ones considered here can 'average away' such errors when building their statistical models. On close inspection, certain parts of the boundary of the lung fields found by PC are in fact judged to be more accurate than the gold standard. This may partly explain the fact that the best PC system achieves better performance for right lung field segmentation than the second observer. Another reason for this fact may be that there are systematic differences between both observers – and the computer algorithms are trained and evaluated with segmentations from the same observer.

There are at least two reasons why the fact that there is no significant difference between PC and the human observer for lung field segmentation does not mean that this segmentation task can be considered 'solved'. First, depending on the usage of the segmentation, the overlap measure Ω may not be a good measure of segmentation performance. Although the overall overlap is excellent, there are certain parts of the lung field, which pose more problems for PC than for a human observer. Second, the JSRT database contained only images of good technical quality, and very few images with gross abnormalities. Such images are much harder to segment for the considered computer methods than for humans, because grossly abnormal cases are usually individually unique and thus not represented in the training set.

Keeping these limitations in mind, let us consider the performance of the different methods for each of the segmentation tasks examined.

Lung Segmentation

Of all objects, the overlap values obtained for the right lung are highest, for both the human observer and all automatic methods. PC obtains better results than the human observer, although the difference is not significant. It is interesting to note how well the shapes produced by PC approximate lung shapes even though no shape information is encoded explicitly in the method.

The left lung is more difficult to segment than the right lung because of the presence of the stomach below the diaphragm that may contain air, and the heart border, which can be difficult to discern. Again, PC is very close to the human observer (no significant difference). There is no indication that any of the methods for lung segmentation proposed in the literature (Section 10.2) achieves segmentation accuracy comparable to human performance.

In most cases, ASM and AAM produce satisfactory results as well, but occasionally left lung segmentation proves problematic. Consider the difficult case on the

right in Figure 10.5, where the border of the enlarged heart is very close to the outer border of the left lung field. Moreover, the heart border is fuzzy, and therefore difficult to locate precisely. ASM followed a different edge and included the heart in the lung field; this can be considered an interpretation error. AAM put the heart border somewhere halfway between the true border and the border followed by ASM, and pushed the border of the lung field outside the rib cage, probably as a result of shape modelling which does not allow the heart border and the lower left lung border to be so close. PC, not hampered by a shape model that cannot deal with this uncommon shape, produces a very satisfying result. Note how the segmentation of the second observer deviates from the gold standard. The second observer probably made an interpretation error in this case. Note also that the Ω values for lungs and heart are similar for this case, although the CTR is very different.

Heart Segmentation

For the heart segmentation, the difference between the second observer and the automatic methods is much larger than for lung segmentation. The agreement between both human observers is much lower as well, the lowest for all objects. The reason for this is that the upper and lower heart border cannot be seen directly on the radiograph. The observers have to infer the location of the heart from the left and right border and anatomical knowledge. The upper heart border is known to be a little below the hilum where the pulmonary arteries enter the lungs. In terms of Ω, PC works best, but ASM and AAM are close. Although AAM has a slightly lower score than ASM on average, its agreement with the gold standard is better for the determination of the CTR.

Clavicle Segmentation

Contrary to the other objects, the clavicles are small. There is no clear difference in performance of any method between left and right clavicle. Clavicle segmentation is a difficult task for various reasons. The bone density can be low, so that the clavicles are hardly visible; there are other, similar edges from ribs in close proximity, and the orientation and position of the clavicles varies enormously. This can be seen from the shape model and the spread of the individual points in Figure 10.2 and 10.3. As a result, segmentation of clavicles is a challenging task. The difference between the computerized methods and the second observer is large.

ASM is the best method for clavicle segmentation, but occasionally there is hardly any overlap between the detected and actual clavicle, as can be seen in the box plots of Figure 10.4 and the most difficult case in Figure 10.5. Running a separate ASM to detect clavicles did not show a clear performance improvement. We believe that the problem of confusing edges and the large variation of clavicle position and orientation are the main reasons for failures with ASM.

AAM performs substantially poorer than ASM for clavicle segmentation, contrary to heart and lung segmentation where the results between the two methods were comparable. This behaviour was anticipated for several reasons. We hypothesize that the dominating factor is the differences in weighting of the clavicles compared to the lung and heart regions. ASM uses independent, equally weighted texture models around each landmark. Thus, the relative 'importance' – wrt. the model-to-image cost function – of each sub object in an ASM is solely determined by its number of landmarks. In this application, each clavicle had approximately half the number of landmarks present in the corresponding lung contour (see Figure 10.1). On the contrary, AAM optimises a global model-to-image fit, accounting for all pixel samples on the object surfaces. This means that no equalization between areas or objects is performed. Consequently, small objects are easily sacrificed for a better fit of large objects. Further, objects with subtle intensity variations – and weakly defined borders – are sacrificed for objects with large intensity variation. Both issues pertain to segmentation of clavicles. In this application the non-globally normalised ASM model has proved desirable for clavicle segmentation, but in other cases such strong priors may lead to problems, typically due to amplification of noise-contaminated signal parts. Using two landmark-based benchmarks, ASM was also shown to outperform AAM on face and brain data in work by Cootes et al. [53].

Secondly, since AAM requires mappings between examples to be homeomorphisms (continuous and invertible), layered objects moving independently will inherently cause problems. In the projection images analysed here, the clavicle position with respect to the lung border is inconsistent. In two examples clavicles were actually above the lungs. Further, the medial endpoints of the clavicles can be inside or outside the lung fields. To obtain a perfect registration between such images an AAM would need to introduce folds, i.e. degenerate piecewise affine warps with inverted mesh normals. Replacing these with thin-plate splines [30] will inevitably also lead to folds. To solve this, objects need to be modelled as independent layers in the texture model. Rogers [185] has previously acknowledged this problem when modelling capillary images.

To assess the practical impact of this on the overlap measure, we have built an AAM for the lung and heart contours only. Warp degeneracy had no noticeable impact on the accuracy of lung and heart localization.

PC undersegments the clavicles, but hardly has any false positives. This can be explained by poor features, which make classification into the class with higher prior probability more likely. A lower threshold for the hard classification could prove helpful, but this has not been investigated in detail. The post-processing of the PC method is counterproductive when the clavicle segmentation produces two or more segments of similar size since only the largest segment is retained. This occurs in the two most difficult cases shown in Figure 10.5.

Determination of CTR

There is a variety of measures that can be computed directly from a segmentation of anatomical structures in a chest radiograph. Such measures can be of great clinical importance and their automatic computation may help in extracting more information from a routine chest examination. In this work, we considered the cardiothoracic ratio. Other possibilities are the area of heart and lungs, the total lung capacity (for which an additional lateral chest film is required) [16], the diaphragm length [21], and the vascular pedicle width (VPW) [91]. Measurement of the diaphragm length and the VPW requires knowledge about the location of certain landmarks in the image, which is known from the point positions obtained by ASM and AAM segmentation.

An automatic system to estimate the CTR was described by Nakamori et al. in [172], in which points along the heart boundary were detected by fitting a Fourier shape to image profiles. This system was used to compute the CTR in 400 radiographs in another study [171] where radiologists had to correct the computer result in 20% of all cases. Automatic determination by the methods presented here is probably substantially more accurate.

Pixel Classification versus Active Shape and Appearance Models

There are some fundamental differences between PC, ASM and AAM. PC does not have a shape model. Therefore it can produce implausible shapes, and is likely to do so when the evidence obtained from image features is not conclusive. This can be observed from the heart borders and the clavicles in Figure 10.5. PC does not require landmarks, only labels per pixel. In that sense it is more general, it can also be applied to tasks where it is difficult or impossible to set corresponding points. PC can also produce a soft classification, which cannot be obtained directly from ASM and AAM. On the other hand, ASM and AAM provide more information than just a binary segmentation; correspondences along contours are established between the search result and any training example. Thus, a registration is obtained and any anatomical landmarks defined by a position on a contour can be inferred on new examples using ASM or AAM.

Another important difference between PC and ASM/AAM is that the latter are based on linear appearance models whereas the PC system uses non-linear features and a non-linear classifier to map appearance and position characteristics to class labels. Had the PC system been restricted to features similar to those used in AAM and ASM (pixel values in the object and along profiles) and a linear classifier, the results would have been much worse.

PC does not employ an iterative optimisation scheme. This avoids the problems typically associated with such schemes, such as ending up in local minima. Although PC is conceptually more simple – and easier to implement – it is computationally more demanding than ASM and AAM. The multi-resolution scheme pro-

posed here, and the approximate kNN classifier make the method usable on these 2D data. With increasing computational power and dedicated optimisations for processing speed, segmentation of 3D data sets with complex PC systems will become routinely feasible as well.

Contrary to both ASM and PC, an AAM also establishes a dense planar correspondence. This dense registration enables that every *interior* point on the model can be localized on an unseen image after AAM search. This combined with a per-pixel statistical model, provides a starting point for e.g. detection of abnormalities in the lung field.

All in all, the choice for a particular segmentation algorithm can be motivated by more than the expected segmentation accuracy, such as computational demands, implementation complexities and the requirements for further analysis.

Improving ASM, AAM and PC

Changes or extensions to the ASM algorithm can address the shape model, the appearance model or the optimisation algorithm. In Table 10.3 it is shown that when the shape model is fitted to the gold standard, the mean overlap remains below the results of PC for the right lung and around the accuracy of the human observer for both lungs. For the clavicles, fitting the shape model leads to an overlap well below that of the human observer. Thus the shape model is not able to capture all shape variations in the test set. It is possible that more examples are needed, or that better results can be obtained with more flexible models. When ASM is started from the gold standard position, the results for all objects are surprisingly close to the actual results of the tuned ASM system. This indicates that the multi-resolution optimisation procedure initialised with the mean shape performs adequately. Still, by comparing the shape model fit to the gold standard and the result of ASM initialised with the gold standard, it can be seen that the solution drifts away from the perfect initialisation. The appearance model is thus not perfect. Non-linear appearance models might lead to better performance [79, 190, 243].

Contrary to ASM, changes to the internal parameters of AAM had little influence on the final result. Therefore there is no default and tuned setting presented for AAM. The extension with whiskers, however, was essential to obtain good performance with AAM. Instead of BFGS, a refinement method based on random sampling should partially avoid the problem of falling into a local minimum. However, this is anticipated to be computationally more demanding due to the high dimensionality of the optimisation space. Table 10.3 also shows that the lung accuracy obtained using AAM with BFGS refinement is very close to the upper bound. This suggests that the cost function hyper surface is indeed very flat and improvements in overlap for the heart and clavicles in Table 10.3 is apparently obtained by starting the AAM search closer to the global minimum in this region.

The good performance for PC depends heavily on the features and the classifier.

Clearly, there are many more feature sets and classifiers that could be evaluated, and feature extraction and selection techniques could be employed. This is an advantage of PC: it formulates segmentation in terms of a standard pattern recognition task. The full vocabulary of techniques from this field can be used.

The post processing stage of PC is simple, but ad hoc. It ensures a single object, without holes and with a somewhat smooth border (due to blurring the posterior probabilities) but the settings of the procedure have not been trained and the optimal settings are unlikely to be the same for all objects. Instead of post processing, one could resort to iterative re-labelling [151], MRF, relaxation labelling. This will almost certainly improve performance, but is substantially more computationally demanding.

Spatial position is an important feature for PC. Without it, the left and right lung fields, for example, would be virtually indistinguishable. The heart segmentation suffers from little image information, and position is a very important feature. The position features are 'raw' positions, however. After the lung fields have been segmented – which can be done very accurately, as has been demonstrated – the spatial position relative to the lung fields could be used instead of the raw (x, y) values. This may improve heart segmentation accuracy. For clavicle segmentation, it may not be too useful, as the position of the clavicles relative to the lung fields varies a lot (Figure 10.3).

Finally, given the fact that there is no single method that outperforms the alternatives among all methods and segmentation tasks considered, it is likely that a hybrid system will achieve superior results. Such a system could combine the results or fuse algorithmic elements of PC, ASM and AAM.

10.7 Conclusion

A large experimental study has been presented in which several versions of three fully automated supervised segmentation algorithms have been compared.

The methods were active shape models (ASM), active appearance models (AAM) and pixel classification (PC). The task was to segment lung fields, heart, and clavicles from standard chest radiographs. Results were evaluated quantitatively, and compared with the performance of an independent human observer. The images, manual annotations and results are available for the research community to facilitate further studies, and so is the AAM implementation [213].

The main conclusions are the following:

1. All methods produce results ranging from excellent to at least fairly accurate for all five segmentation tasks considered, using the same settings for all objects. This demonstrates the versatility and flexibility of general supervised

image segmentation methods.

2. The cardiothoracic ratio can be determined automatically with high accuracy. Automatic CTR determination could be provided in a clinical workstation.

3. The best method for lung field segmentation is PC, followed by ASM, AAM (differences between ASM and AAM not significant); for heart segmentation AAM, followed by PC, ASM (only a significant difference between AAM and ASM); for clavicle segmentation ASM, followed by AAM, PC (all differences significant except for those between AAM and PC for the right clavicle).

4. For segmentation of both right and left lung, there is no significant difference between manual segmentation by an independent human observer and PC. In all other cases, the human observer performed significantly better than the best computer system. For clavicles, the differences between computerized methods and a human observer were largest. This indicates that accurate computerized segmentation of clavicles and heart in chest radiographs is still an open problem.

5. Significant performance improvements for all five segmentation tasks were obtained by tuning the parameters of ASM and by including information from outside the object in the AAM. Refining the AAM result with a general-purpose optimisation method improved performance for all objects except clavicles. This shows that one must pay careful attention to parameter settings and variations on the basic algorithms that have been proposed in the literature when applying ASM or AAM in practice.

6. The output of pixel classification can lead to ragged borders. A simple post-processing stage decreases the error significantly for lung fields and heart segmentation. The presented PC method obtains excellent results in these tasks, despite its lack of a shape model.

Acknowledgements

The authors gratefully acknowledge R. Nievelstein for supervising the manual segmentations and G. Mochel and A. Scheenstra for each clicking around 75,000 points to segment the images.

CHAPTER 11

Unsupervised Motion-compensation of Multi-slice Cardiac Perfusion MRI

Mikkel B. Stegmann, Hildur Ólafsdóttir, and Henrik B. W. Larsson

Abstract

This paper presents a novel method for registration of single and multi-slice cardiac perfusion MRI. Utilising off-line computer intensive analyses of variance and clustering in an annotated training set, the presented method is capable of providing registration without any manual interaction in less than a second per frame. Changes in image intensity during the bolus passage are modelled by a slice-coupled active appearance model, which is augmented with a cluster analysis of the training set. Landmark correspondences are optimised using the MDL framework due to Davies et al. Image search is verified and stabilised using perfusion specific prior models of pose and shape estimated from training data. Qualitative and quantitative validation of the method is carried out using 2000 clinical quality, short-axis, perfusion MR slice images, acquired from ten freely breathing patients with acute myocardial infarction. Despite evident perfusion deficits and varying image quality in the limited training set, a leave-one-out cross validation of the method showed a mean point to curve distance of 1.25 ± 0.36 pixels for the left and right ventricle combined. We conclude that this learning-based method holds great promise for the automation of cardiac perfusion investigations, due to its accuracy, robustness and generalisation ability.

Keywords: motion-compensation, registration, cardiac perfusion MRI, active appearance models, minimum description length

11.1 Introduction

Within the last decade magnetic resonance imaging (MRI) has been proven able to assess myocardial perfusion in an accurate and safe manner, see e.g. [143]. While scanning times have improved substantially, the amount of manual post-processing required remains to render the method prohibitive to clinical practice. Marking up points of correspondence on the myocardium constitutes a major part of this manual labour, which is essential to ensure compensation of motion during the perfusion sequence. This paper presents a novel approach aiming at replacing this resource demanding, tedious and error prone task with an automated image analysis method, which provides a structured way of collecting and applying expert knowledge given by medical doctors into a learning-based framework.

Equal to most problems within medical image analysis, registration of perfusion MRI requires a very robust technique. This is partly due to the many factors during the MR image acquisition that deteriorate image quality such as resolution/time trade-offs, field inhomogeneity, susceptibility artefacts, flow artefacts, motion artefacts, partial volume effects, chemical shift artefacts, et cetera. Further, robustness is also required towards the severe biological inter-subject variation as well as variations in scan planning and scanner configuration. Finally, to be clinically viable, a registration technique should be able to deal with pathologies – which in this case mostly include perfusion deficits – without sacrificing accuracy or reliability.

In addition to the above requirements, we believe that methods – insofar possible – should be driven by the characteristics of the data, opposed to having explicitly encoded constraints regarding the properties of the domain problem, e.g. size, position, intensity, et cetera. Consequently, we have based our method on the learning-based Active Appearance Models (AAMs) framework [52, 88]. In this paper we analyse the differences between the single-image oriented AAMs and time-series of multi-slice perfusion MRI and propose a framework designed to meet the stated requirements.

The clinical goal of this work is to automate the generation of perfusion maps[1] and ultimately; the actual perfusion measured in ml/(g·min), see e.g. [143]. However, our method is currently validated using the registration accuracy, since this constitutes the core requirement for obtaining accurately motion compensated perfusion estimates.

The paper is organised as follows. A brief review of the existing body of work pertinent to image analysis of perfusion MRI is given in Section 11.2. Section 11.3 describes the data used for this study. Section 11.4 treats the methodological backgrounds and describes the proposed framework. Section 11.5 presents a qualitative and quantitative validation. Section 11.6 provides a discussion of the obtained results, the merits and demerits of the method, and a view to future work. Finally,

[1] These include contrast enhancement ratio, transit time, maximum upslope, upslope integral and time-to-peak, et cetera. See e.g. [40, 42, 124, 168].

Section 11.7 draws some concluding remarks.

This paper represents a direct extension of the work reported in [224].

11.2 Related Work

In this section, the current status of registration of MR cardiac perfusion images will be considered. Refer to Table 11.1 for an overview of the methods stated below.

Note that all the following approaches register one slice sequence at a time independent of the remaining slices from the scan. Therefore computation times apply to a single slice sequence, unlike what is presented in this work.

Yang et al. [254] use the phase difference of the raw MR data between successive image frames to correct for translational motion. Shape changes are compensated for by using a deformable model. They show example results from one patient but lack further validation of the method.

Behloul et al. [18] simplify the problem of myocardial boundary detection by adding a slice from functional MRI to the perfusion sequence. They segment the functional slice by fuzzy clustering and fuzzy inference systems and warp the result to the perfusion sequence. The method requires manual definition of a rectangular region of interest (ROI) on each image of the perfusion sequence. The authors claim that their method can be applied in real-time but unfortunately it lacks validation.

Bidaut and Vallee [23] use a multi resolution translation/rotation based registration minimising the mean squared differences (MSD) between each image frame and a reference frame. They evaluate their method using 8 sequences of 90 frames from 8 patients with stable coronary artery disease (CAD). The validation is based on a comparison of image correlation factors in addition to motion reduction measurements for manual (ground truth) and automatic registration. Their computation time is 5 minutes per 100 time frames.

Breeuwer et al. [41], Breeuwer et al. [42], Spreeuwers and Breeuwer [200] also use a translation/rotation based registration but with normalised cross-correlation as a similarity measure. After the registration, an exact detection of endocardial and RV boundaries is obtained by region growing on feature images. The epicardial boundaries are detected by a deformable snake model. In [41], [42], the method is evaluated on 36 perfusion sequences from 12 patients by comparing the calculated perfusion parameters to estimated values from X-ray angiograms of the same patients. They obtain good correspondence for 11 of the patients. Their computation time is less than 4 minutes per slice sequence. Spreeuwers and Breeuwer [200] evaluate qualitatively on 30 perfusion sequences from 14 scans. They succeed in 26 out of 30 sequences. Their computation time is 25 seconds per 70 time frames.

Gallippi and Gregg [102] present a statistics based registration method using deformable template matching. The similarity measure depends on local brightness

variations and edge directions. They evaluate their method on 12 patients and measure its quality by means of motion reduction and comparison of signal-intensity (SI) curves prior to and after registration.

Ablitt et al. [1] use tissue tagging to correct for through-plane distortion on a multi-slice sequence. The in-plane motion in each slice is corrected for using free-form image registration with partial least squares regression (PLSR) deformation learning. They evaluate their method on a set of 8 patients with CAD and 5 healthy individuals in addition to a synthetic data set. The in-vivo data sets include 3 slices with 50 frames per slice. The accuracy of the method is measured relatively by comparing SI curves before and after registration. Gao et al. [103] do further studies on the PLSR deformation modelling. The method is evaluated using 9 patients with CAD, each including 3 slices and 50 frames per slice, as well as a synthetic dataset. The accuracy of the method is measured by comparing motion reduction for a free-form image registration versus the proposed method.

Gupta et al. [109], Bansal and Funka-Lea [15] and Bracoud et al. [40] use translation based registration methods. Gupta et al. with cross-correlation and the latter two with mutual information (MI) as a similarity measure. None of the methods compensate for rotation or deformable shape changes. Gupta et al. validate on 10 patients, measuring motion reduction and comparing SI curves. The computation time of 50 frames sequence is 5 seconds but 30 out of 433 frames needed manual post processing. Bracoud et al. evaluate their method on 5 patients (3 slices, 30 frames per sequence) with stable CAD. They compare histograms before and after registration, inspect SI curves and perfusion maps and measure motion reduction. Their method takes 8 minutes for a 30 frame sequence including a manual definition of a rectangular ROI in one image of the sequence.

Spreeuwers et al. [201] present a refinement method to optimally place the myocardial borders after manually (or automatically) tracing them. They investigate the effect of displacements of the myocardium in perfusion MRI and use that information to correct the boundaries. They apply their method on 9 image sequences on which the myocardium has been traced, and compare perfusion parameters before and after correction.

Table 11.1: Overview of registration methods for MR cardiac perfusion images.

Ref.	Method	Similarity measure	Data	Validation method	Breathing	Computation time	Fully automated
[254]	Translation based registration, deformable model	-	1 patient (100 frames)	SI curves	-	-	No
[18]	Segmentation of functional MRI mapped to perf. seq.	-	-	-	-	Real-time	No
[23]	Translation/rotation based registration	MSD	8 patients (8 sequences, 90 fr.)	Motion reduction, image correlation factors	-	5 min per 100 fr.	Yes
[41], [42]	Translation/rotation based registration, region growing, snake model	Normalised cross correlation	12 patients (36 seq., 60 fr.)	Quantitative comparison of perf. parameters	No	4 min per seq.	Yes
[200]	Translation/rotation based registration, region growing, snake model	Normalised cross correlation	14 (30 seq., 70 fr.)	Visual inspection	No	25 sec per 70 fr.	Yes
[102]	Deformable template matching	Brightness, edge directions	12 patients	Motion reduction, SI curves	Yes	-	Yes
[15]	Translation based template matching using edge and intensity information	Mutual information	-	-	-	-	Yes
[1]	Free-form image registration with PLSR deformation model	Cross correlation	8 patients (24 seq., 50 fr.), 5 healthy (15 seq., 50 fr.), synthetic data	SI curves	Yes	-	Yes
[103]	PLSR Deformation model	-	9 patients (27 seq., 50 fr.), synthetic data	Motion reduction	Yes	-	Yes
[109]	Translation based registration	Cross correlation	10 patients (4-6 slices, 433 frames in total)	Motion reduction, SI curves	No/yes[a]	5 sec per 50 fr.	No
[40]	Translation based registration	Mutual Information	5 patients (15 seq., 30 fr.)	Histograms, SI curves, perf. maps, motion reduction	Yes	8 min per 30 fr.	No
[201]	Sensitivity analysis on perf. assessment	-	9 seq.	Comparison of perf. parameters	-	-	-

[a] Breath-holding while comfortable, free-breathing after that.

11.3 Data Material

The data material comprises 2500 myocardial perfusion, short-axis, magnetic resonance images (MRI) obtained from ten freely breathing patients with acute myocardial infarction. Five slices of the first 50 sequential frame images each were selected before, during and after the bolus of contrast. The total sequence length acquired was approximately 70 frames. This study is performed using the 2000 images of the latter four slices of this set. The contrast agent was gadolinium diethylenetriaminopentaacetic acid (Gd-DTPA). Registration relative to the heart-cycle was obtained using ECG-triggered acquisition from a whole-body MR unit, Siemens Vision, operating at 1.5 Tesla. Slice acquisition was restarted at every third R-peak, thus providing an approximate frame time of three seconds, depending on the heart rate. MR pulse sequence: inversion recovery turbo-FLASH (fast low-angle shot), matrix size=128×128, field of view=300×300 mm, slice thickness=10 mm, inter slice gap=0 mm, storage bit depth=16 bit, inversion time, TI=598 ms, 209 ms, 792 ms, 404 ms. See Figure 11.1 for an example multi-slice frame from the first patient showing four slices in anatomical order (in basal-apex direction from left to right). All slice images in the following sections will also be shown as a row of single images. The left digit in each slice shown in Figure 11.1 denotes the frame number. The right digit denotes the order in which each slice was acquired (see the inversion times above).

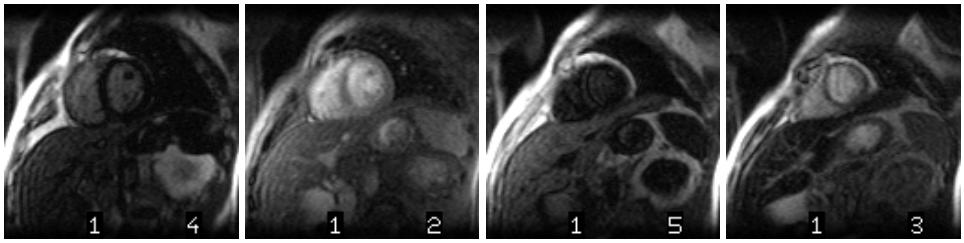


Figure 11.1: Four slice images of patient 1 prior to bolus arrival (frame 1). Inversion time, TI=598 ms, 209 ms, 792 ms, 404 ms.

11.4 Methods

11.4.1 Myocardial Perfusion Imaging

Developments in MR-technology during the past decade have made it possible to acquire physiological information about dynamic processes in the human body. As an example of such, myocardial perfusion imaging encompasses assessment of myocardial perfusion at rest and during stress (e.g. pharmacological). By injecting a bolus of a paramagnetic contrast substance the myocardial perfusion mechanism

can be quantified. Thus, perfusion MR qualifies as an essential instrument in the assessment of ischemic heart diseases. As the contrast agent tags the blood stream and amplifies the MR signal by a shortening of the T_1 relaxation time, areas of the myocardium served by diseased arteries show a delayed and attenuated response. Acquisition is carried out dynamically and registered to the heart cycle using ECG-triggering. Images are typically acquired from one or more short-axis slices every n-th heartbeat, where through-plane resolution is traded for temporal resolution and vice versa. If the acquisition time-window is sufficiently short (typically < 40 seconds), breath-hold can be used to remove respiration artefacts.

We further mention that imperfect ECG-triggering can give rise to depiction of erroneous heart-phases, in which tissue correspondences can be partially shattered, due to the long-axis movement of the left ventricle during the heart cycle.

In summary, the most prominent sources of variation in shape and appearance found in perfusion MRI – aside from general MR acquisition artefacts – include: biological inter-subject variation, scan planning, contrast passage, contrast uptake, respiration and inaccuracies in ECG-triggering. Additionally, when images are analysed on the basis of manual annotations, this mark-up process introduces inter- and intra-subject variability.

11.4.2 Active Appearance Models

Active Appearance Models (AAMs) [52, 88] were introduced as a learning-based method for registration and interpretation of face images. These models are built based on a set of order-less annotated images. By being a generic approach for image registration and image interpretation, medical applications were soon to follow. Refer to [213] for a summary of medical AAM applications.

Formally, AAMs establish a compact parameterisation of object variability, as learned from a representative training set. The modelled object properties are usually shape and pixel intensities. The latter is henceforward denoted *texture*. From these quantities new images similar to the training set can be generated. Objects are defined by marking up each example with points of correspondence (i.e. landmarks) over the training set either by hand, or by semi- to completely automated methods (see e.g. [78]). Using a learning-based optimisation strategy, AAMs can be rapidly fitted to unseen images, thus providing image registration and interpretation.

Variability is modelled by means of a Principal Component Analysis (PCA), i.e. an eigenanalysis of the dispersions of shape and texture. Let there be given Q training examples for an object class, and let each example be represented by a set of N landmark points and M texture samples. Shape examples are aligned to a normalised common mean using a Generalised Procrustes Analysis, [108]. Texture examples are warped into correspondence using a piece-wise affine warp, normalised, and subsequently sampled from this *shape-free* reference. Typically, this geometrical reference frame is the Procrustes mean shape. Let $\mathbf{s}$ and $\mathbf{t}$ denote a synthesised shape

and texture and let $\bar{\mathbf{s}}$ and $\bar{\mathbf{t}}$ denote the corresponding sample means. New instances are now generated by adjusting the principal component scores, $\mathbf{b}_s$ and $\mathbf{b}_t$ in

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{\Phi}_s \mathbf{b}_s \quad , \quad \mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_t \mathbf{b}_t \tag{11.1}$$

where $\mathbf{\Phi}_s$ and $\mathbf{\Phi}_t$ are eigenvectors of the shape and texture dispersions estimated from the training set. To obtain a combined shape and texture parameterisation, $\mathbf{c}$, the values of $\mathbf{b}_s$ and $\mathbf{b}_t$ over the training set are combined into

$$\mathbf{b} = \begin{bmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_t \end{bmatrix} = \begin{bmatrix} \mathbf{W}_s \mathbf{\Phi}_s^{\mathrm{T}} (\mathbf{s} - \bar{\mathbf{s}}) \\ \mathbf{\Phi}_t^{\mathrm{T}} (\mathbf{t} - \bar{\mathbf{t}}) \end{bmatrix}. \tag{11.2}$$

A suitable weighting between pixel distances and pixel intensities is carried out through the diagonal matrix $\mathbf{W}_s$. To recover any correlation between shape and texture the two eigenspaces are usually coupled through a third PC transform,

$$\mathbf{b} = \mathbf{\Phi}_c \mathbf{c} = \begin{bmatrix} \mathbf{\Phi}_{c,s} \\ \mathbf{\Phi}_{c,t} \end{bmatrix} \mathbf{c}, \tag{11.3}$$

obtaining the combined appearance model parameters, $\mathbf{c}$, that generate new object instances by

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{\Phi}_s \mathbf{W}_s^{-1} \mathbf{\Phi}_{c,s} \mathbf{c} \quad , \quad \mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_t \mathbf{\Phi}_{c,t} \mathbf{c}. \tag{11.4}$$

The object instance, $\{\mathbf{s}, \mathbf{t}\}$, is synthesised into an image by warping the pixel intensities of $\mathbf{t}$ into the geometry of the shape $\mathbf{s}$ and applying the current pose parameters $\mathbf{p} = [\ t_x\ t_y\ s\ \theta\ ]^{\mathsf{T}}$ where t_x, t_y and θ denotes in-plane translation and rotation, and s denotes the shape size.

Using a least-squares criterion the model is matched to an unseen image by an iterative updating scheme based on a fixed Jacobian estimate [54] or a principal component regression [52]. For this work we have used the former approach, treated in further detail in [213]. In brief, this allows model parameters, $\{\mathbf{c}, \mathbf{p}\}$, to be updated based on the difference between the model texture, $\mathbf{t}$, and image texture, $\mathbf{t}_{image}$, the latter being the image texture covered by the model. Let $\mathbf{R}$ denote a fixed update matrix pre-calculated from the training set. Updates are then estimated by

$$[\ \delta\mathbf{c}^{\mathsf{T}}\ \delta\mathbf{p}^{\mathsf{T}}\ ]^{\mathsf{T}} = \mathbf{R}(\mathbf{t}_{image} - \mathbf{t}). \tag{11.5}$$

For further details on AAMs refer to [52, 54, 65].

In our application we mention that due to the inherent representation of texture vectors in AAMs, pose- and shape-compensated images are *directly* obtained by projecting each texture vector into the shape-free reference frame.

Thereby a per-pixel correspondence over the complete perfusion sequence can be obtained, ready to be analysed by a perfusion map, or by a perfusion model, see e.g. [143].

11.4.3 Shape Annotation and Landmark Correspondences

One important step towards a robust shape model is to generate a training set of shapes with good landmark correspondence. A semi-automatic procedure involves two steps: i) Extraction of shape contours from the images. ii) Definition of marks at corresponding locations across the set of contours.

The first step is a manual procedure while the landmark correspondences can be obtained automatically using Minimum Description Length (MDL) shape modelling introduced by Davies et al. [76].

An annotation tool has been developed in order to achieve the first step, extraction of the shape contours. The tool enables the user to extract the desired object by marking points on the contour using a mouse or a digitiser. We use a WACOM tablet digitiser, which both gives faster and more accurate annotation than using a mouse. The cardiac shape includes two anatomical landmarks, the points where the RV border meets the epicardial, called anterior junction and inferior junction as shown in Figure 11.2.

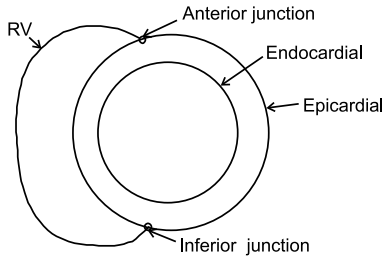


Figure 11.2: Anatomical landmarks on the cardiac shape: Anterior junction and inferior junction between right and left ventricle.

The anatomical landmarks are included in the manual process, i.e. the user defines those while annotating. After placing points on the object, the contour itself is given by an interpolating cubic spline between the points.

After this manual step of the procedure, MDL optimisation gives a reparameterisation, representing an optimal point correspondence, of the given shape contours. The MDL approach is based on minimising the description length, a quantity adopted from information theory. The description length in this case is the cost of transmitting the PCA coded model parameters in addition to the transmission cost of the encoded data values. This means that MDL balances the complexity of the model against how well the model fits the data. For further details of MDL shape modelling see [75] and [77].

We have used the freely available MDL implementation from [235]. Roughly, the approach implemented here is to slide points on one shape at a time according to an adaptive step length, align the set of shapes, do PCA of the set and calculate the

description length. For each move, the description length is compared to the current best one. Based on comparison, this move is either accepted or rejected.

MDL shape modelling cannot be applied directly to the cardiac shape, since it is not possible to parameterise it. Furthermore, the fixed points (here, the anatomical landmarks) can give rise to problems since they are not end points. The approach used here was to split the shape into five open fixed-end contours. The epicardial contour is split at the anatomical landmarks and the endocardial contour is split at pseudo landmarks placed at minimum distance from the anatomical landmarks. This is clarified in Figure 11.3.

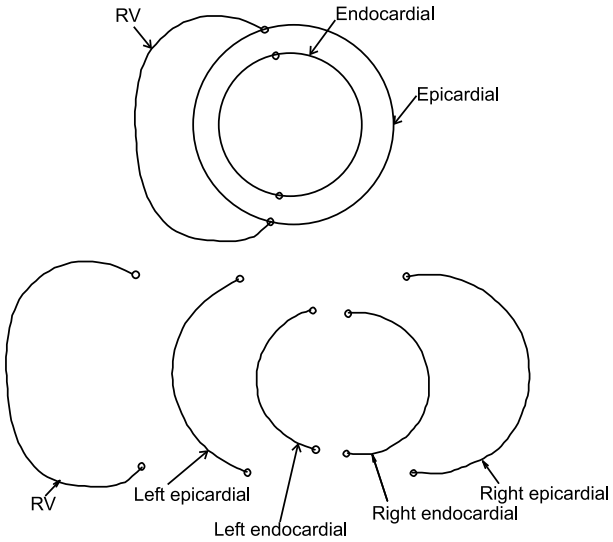


Figure 11.3: The cardiac shape split into five open fixed-end contours. The MDL approach is applied to each of the five contour sets.

The MDL approach is applied to each of the five contour sets. The aim is to optimise each contour set locally, but with influence from the full shape. This is done by aligning the full shape (all five contour sets). The PCA and the optimisation itself, i.e. the sliding of points, is performed on the local contour set only.

With the large data set of 2000 shapes, the MDL optimisation process becomes very slow. By noticing that the most shape variation occurs between patients and between slices, the approach used here is to select the last frame of each slice sequence for the MDL optimisation. This reduced the set to 40 shapes (for each contour type) which leads to an acceptable computation time. The output from MDL is a set of parameterisation functions; one for each contour type for each of the shapes. These parameterisation functions are applied to the remaining frames of the sequence. Eventually, the five contour sets are combined again to form the full training set.

11.4.4 Modelling of Perfusion MRI Time-series

Since perfusion MRI sequences differ in structure from the single-image oriented AAMs, this section will discuss the issues of data modelling.

First, one should recognise that the signal variation of the left ventricle (LV) and the right ventricle (RV) is very small prior to contrast arrival. Due to this lack of image contrast, the standardisation of texture vectors normally used in AAMs would result in severe amplification of scanner noise. Hence, only the texture mean is removed in the following pre-processing of image texture vectors.

Deviations from the assumptions in standard AAMs are not only encountered with respect to texture normalisations. The relationship between shape and texture also differs. It is safe to assume that the process generating shape variation remains stationary throughout a sequence, contrary to the highly non-stationary texture process. Consequently, shape and texture are treated independently by removing the combined PCA, i.e. $\mathbf{\Phi}_c = \mathbf{I}$. Treating each perfusion sequence as one observation (as done on cardiac cine MRI in [159]) is not feasible due to the random fluctuations in pose and shape induced by the variation sources mentioned in Section 11.4.1; whereof respiration dominates.

However, joint modelling of texture – decoupled from changes in shape and pose – for a complete sequence is possible by introducing an annotation of the temporal relationship of each sequence frame. This may not be desirable, as only a few proper temporal landmarks exists (during the actual bolus passage). Further, building sequence models would require far more training examples to model this behaviour properly, compared to a simpler and less constrained frame-based model. Consequently, given the low number of subjects, we will treat each frame in a sequence as an observation.

Circumventing the need for large training sets unfortunately violates a basic assumption in AAMs, namely that the variation in texture is well modelled by a single multivariate Gaussian model. Due to the radical changes in intensity during contrast passage and uptake this is clearly not the case. On a coarse level we can split the sequence into pre-contrast arrival, contrast agent entering the RV, LV, and the myocardium.

Since the texture parameters of the texture model shown in Equation 11.1 are ranked by their variance over the training set, this alleged clustered behaviour should be confirmed by visual inspection of a few highly-ranked parameters from $\mathbf{b}_t$. Figure 11.4 shows the three most significant texture parameters, $\{b_{t,i}\}_1^3$, from a texture model built from the ten available perfusion sequences. This figure exhibits a pronounced clustering and thus verifies our concerns. Modelling this multimodal distribution with a multivariate Gaussian is clearly not justified and gives rise to several problems. Most problematic is that the resulting model is not specific to the given problem. Thus, it can easily generate textures that are not plausible to occur during a perfusion bolus passage. This would in turn lead to potential false posi-

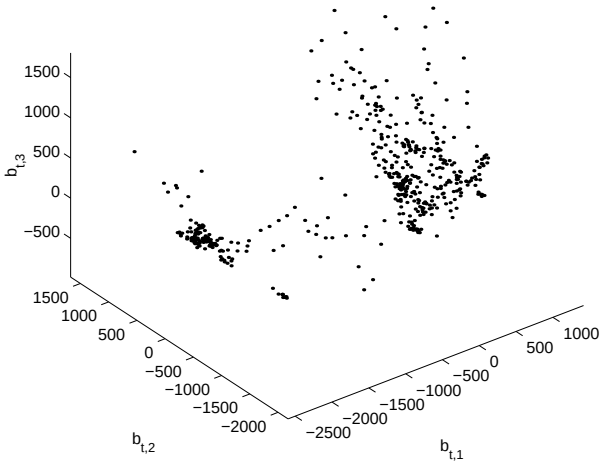


Figure 11.4: First, second and third principal component of 500 texture vectors from ten perfusion sequences.

tives during the model-to-image matching and will be dealt with in the following.

11.4.5 Adding Cluster Awareness

To model the distribution of textures we propose an unsupervised learning approach that models texture variation using an ensemble of linear subspaces in lieu of the unimodal linear model employed in AAMs. Alternatively to clustering, these subspaces could have been given by an operator, which identifies different phases of each perfusion sequence. However, to reduce, i) the tedious burden of training set generation, and ii) inter- and intra-observer variability, supervised learning was rejected. Further, we like to leave the option open to evaluate different ensemble sizes in the future when more training data becomes available. This will obviously be very tedious in the case of manual labelling.

Although the machine learning literature offers an abundance of clustering methods, it is generally agreed upon that no *silver bullet* exists. We have chosen a k-means clustering [97] combined with a Monte Carlo simulation scheme where several clusterings are carried out, based on different initial random class centres. The final clustering is chosen using a minimax criterion, where the clustering having the smallest maximum distance to the nearest class centre is chosen.

The obtained clustering using five classes ($k = 5$) of the data set is shown in Figure 11.5. From this clustering, a set of linear texture subspaces, $\{\mathbf{\Phi}_{t,i}\}_{i=1}^{k}$, is ob-

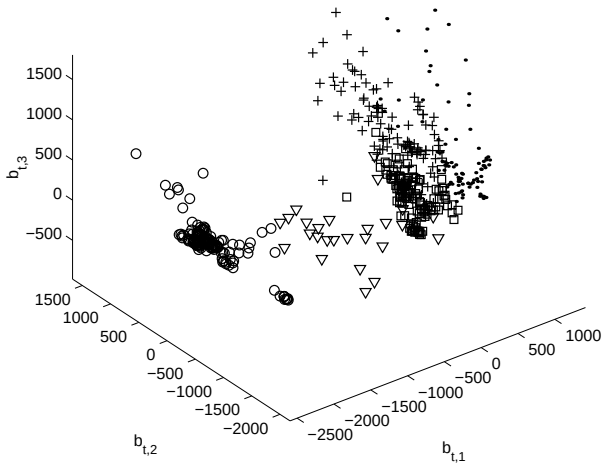


Figure 11.5: First, second and third principal component of 500 texture vectors from ten perfusion sequences clustered into five classes.

tained directly by k separate texture PCAs. A corresponding set of model parameter update matrices, $\{\mathbf{R}_i\}_{i=1}^{k}$, is obtained following the procedure in [54] using a displacement scheme specified in [213].

As texture changes in a sequence are deemed to be unrelated to shape changes, building a joint shape model, $\mathbf{\Phi}_s$, from all frames in all sequences will yield the best estimate of inter- and intra-subject shape variability. We denote this composition of a single shape model and an ensemble of texture models with associated parameter update matrices for a *Cluster-aware AAM* (CAAM). A CAAM can thus be applied to any domain problem that justifies uncoupled shape and texture eigenspaces and requires piecewise-linear modelling of texture variability.

Fitting a CAAM to unseen images now involves choosing the appropriate texture subspace. As a reasonable choice for k is typically fairly limited, model selection is here performed by exhaustively trying all models and selecting the model producing the best model-to-image fit, subject to a set of constraints given later in this paper. To increase performance during model fitting, model selection could be accomplished by a classification of the texture vector into the set of training classes. However, this has not been tried in the current work.

To choose k prior knowledge can be employed. However, being an optimisation problem in one positive integer variable, a data-driven method should be preferred where k is estimated using cross-validation on the training set. However, observe that the number of available training examples for model estimation within a single class is inversely proportional with k. Due to this fact, we have fixed k to five

classes for all experiments as a reasonable compromise, since conclusions regarding the optimal number of classes would remain fragile, due to the limited number of training subjects available.

11.4.6 Modelling Object Interfaces

To add notion about the interface between the heart and the surrounding tissue et cetera, we include intensity samples in the proximity of the heart into the texture model.

This is carried out by sampling landmark normals relative to the current shape size, similar to the approach by Active Shape Models, [68]. We denote these normals *whiskers*. In our application, this will include the LV/lung interface and LV-RV/abdomen interface into the texture models and thus require these to be present in the unseen image to provide a good model-to-image fit. We have weighted whisker samples so that they constitute one third of the texture variation in all of the subsequent experiments.

This approach has earlier been shown to have a positive, and significant, impact on the registration accuracy, in a case study on corpus callosum registration in brain MRI [212].

11.4.7 Multi-slice Shape Modelling

Up to this point an observation has been associated to a single image. However, cardiac perfusion MRI is typically acquired as multi-slice images in a spatial arrangement. This enables regional assessment of the perfusion mechanism both in plane as well as through plane. Modelling such slice images independently will per se disregard any spatial and/or intensity coherence between slices. Conversely, a joint modelling of each set of slices will provide a constrained basis, in which slices with well-defined appearance will restrict slices with diffuse appearance from diverging.

Coupled AAMs were first explored by Cootes et al. [71] for interpretation of multi-view face images. Later Mitchell et al. [159] coupled time-sequences of cardiac MRI, an approach which was also pursued in [33, 36, 37, 146, 198]. Recently, long and short axis MRI, and end-diastolic and end-systolic angiograms were fused by a similar approach by Lelieveldt et al. [147] and Oost et al. [175], respectively.

How coupled AAMs should be constructed is application specific. For our application the following approach was taken. Let a shape consisting of N points in two dimensions be defined by a $2N$ vector,

$$\mathbf{s} = [\, x_1 \; y_1 \; x_2 \; y_2 \; \ldots \; x_N \; y_N \,]^{\mathsf{T}}. \tag{11.6}$$

A multi-slice observation consisting of a stack of C slices is then composed of the set of shapes; $\{\mathbf{s}_i\}_{i=1}^{C}$ by concatenation of the unmodified slice image shape coordinates

prior to normalisation et cetera,

$$\mathbf{s}_{frame} = \left[\begin{array}{cccc} \mathbf{s}_1^\mathsf{T} & \mathbf{s}_2^\mathsf{T} & \ldots & \mathbf{s}_C^\mathsf{T} \end{array} \right]^\mathsf{T}. \tag{11.7}$$

Any subsequent shape analysis is left unchanged. Image sampling is carried out using the appropriate slice for each related subpart of the combined shape vector. This produces a model where inter-slice pose relations are modelled by the shape changes, which is clearly desirable for our application. Should inter-slice differences in pose not be constrained; then slice concatenation should be applied to the Procrustes tangent space shape coordinates instead. This will thus introduce the need for $(C-1) \times 4$ extra pose parameters into the model (in the 2D case).

11.4.8 Estimating and Enforcing Pose and Shape Priors

The fact that changes in pose and shape are unrelated to the change of texture is highly useful for initialising and constraining the model fitting process in each frame. Further, it can validate the final registration results as well as the intermediate results for each texture class. Consequently, if it is possible to obtain reliable estimates of the shape and pose variation in a subpart of the sequence these can be used as constraints in the remaining part of that sequence. This is the case in the latter part of a bolus passage where the contrast agent has been washed out of the RV and LV, only leaving the subtle changes stemming from the perfusion mechanism in the myocardium. Hence, we propose to estimate prior distributions of pose and shape from the latter part of a perfusion sequence of P frames.

Let κ, γ, D_{max} denote a set of user-selectable, dimensionless, constants controlling the influence of the priors. Then, let $\mathbf{\Sigma}$ denote the dispersion matrix of the sequence pose parameters and let σ denote the standard deviations of the sequence shape parameters. How these two quantities are estimated is treated in Appendix 11.A. Further, let $\mathbf{F}_t$ denote the t-th frame. Let the set of frames $\{\mathbf{F}_t\}_{t=1}^{S-1}$ denote the unstable period, and the frames $\{\mathbf{F}_t\}_{t=S}^{P}$ denote the stable period. A formal scheme for exploiting these priors can then be formulated as shown in Algorithm 11.1.

During CAAM search in the unstable period, pose and shape priors are used to stabilise parameter updates by limiting the maximal update step. To simplify notation, the time index t is omitted from this point on. Let Σ_{ij} denote the element in the i-th row and j-th column of $\mathbf{\Sigma}$, and let p_i denote the i-th element of $\mathbf{p}$. Pose parameter updates can now be constrained using the following simple clamping approach:

$$\delta p_i = \begin{cases} \text{sign}(\delta p_i)\kappa\sqrt{\Sigma_{ii}} & \text{if } |\delta p_i| > \kappa\sqrt{\Sigma_{ii}} \\ \delta p_j & \text{otherwise.} \end{cases} \tag{11.8}$$

The constant κ acts thus as a clamping constant given in units of standard deviations of pose variation as estimated from the training sequences. In all experiments we have used $\kappa = 0.5$. As an alternative to this hyper cuboid constraint we could use a projection onto the hyper ellipsoid given by an eigenanalysis of $\mathbf{\Sigma}$.

Algorithm 11.1 Sequence prior augmented CAAM search

Require: S, κ, γ, D_{max}, $\mathbf{\Sigma}$ and σ
1: $\overline{\mathbf{p}}$ = initialisation
2: $\overline{\mathbf{b}}_s$ = initialisation
3: **for** $t = P$ down to S **do**
4: $\{\mathbf{p}_t, \mathbf{b}_{s,t}\}$ = CAAM search started at $\{\overline{\mathbf{p}}, \overline{\mathbf{b}}_s\}$ in $\mathbf{F}_t$
5: $\overline{\mathbf{p}} = \frac{1}{P-t+1} \sum_{j=t}^{P} \mathbf{p}_j$
6: $\overline{\mathbf{b}}_s = \frac{1}{P-t+1} \sum_{j=t}^{P} \mathbf{b}_{s,j}$
7: **end for**
8: **for** $t = S - 1$ down to 1 **do**
9: $\{\mathbf{p}_t, \mathbf{b}_{s,t}\}$ = Constrained CAAM search started at $\{\overline{\mathbf{p}}, \overline{\mathbf{b}}_s\}$ in $\mathbf{F}_t$
(constrain using κ, γ, D_{max}, $\mathbf{\Sigma}$, $\overline{\mathbf{b}}_s$, σ)
10: **end for**

Likewise, we also exploit the prior knowledge of sequence shape variation, σ, (as obtained from the training set) in a similar clamping approach,

$$b_{s,i} = \begin{cases} \overline{b}_{s,i} + \text{sign}(b_{s,i} - \overline{b}_{s,i})\gamma\sigma_i & \text{if } |b_{s,i} - \overline{b}_{s,i}| > \gamma\sigma_i \\ b_{s,i} & \text{otherwise,} \end{cases} \tag{11.9}$$

where γ denotes the maximally accepted distance from the mean in units of standard deviations. We have used $\gamma = 3$ in our experiments.

After ended CAAM search in the unstable period, pose prior enforcement is determined by testing the Mahalanobis distance to the pose distribution:

$$\{\mathbf{p}, \mathbf{b}_s\} = \begin{cases} \{\overline{\mathbf{p}}, \overline{\mathbf{b}_s}\} & \text{if } D_{max}^2 < (\mathbf{p} - \overline{\mathbf{p}})^\mathsf{T} \mathbf{\Sigma}^{-1} (\mathbf{p} - \overline{\mathbf{p}}) \\ \{\mathbf{p}, \mathbf{b}_s\} & \text{otherwise} \end{cases} \tag{11.10}$$

Under the assumption of normally distributed pose parameters, we have chosen to use $D_{max} = 3$. Hence, implausible pose solutions are discarded and replaced with the maximum likelihood of the prior; the mean configuration. Such solutions are flagged and the system operator can be prompted for manual assistance. This pose prior enforcement is also employed during texture model selection so that solutions failing the above condition are excluded from the competition of being the selected for the current frame. This step is essential, since highly deviating pose configurations easily can produce model-to-image fits that will outperform the true positives.

11.4.9 Model Initialisation

The initial pose, shape and texture parameter configuration of the CAAM, in frame P, is determined by exploiting the convergence radius of the AAM search. Within this radius the AAM will converge to a plausible fit. Parameters with variation outside this radius over the training set are semi-exhaustively searched, with grid

spacing less than twice the radius, by starting the AAM search at each grid point. Each search result represents a potential initialisation candidate. To add robustness this is implemented in a candidate scheme, where several candidates compete in an evolutionary process to become the initial configuration. See [205] for the details of this algorithm.

Similar to the normal CAAM search, initialisation is performed exhaustively for all texture models and subsequently choosing the model providing the best fit.

11.5 Experimental Results

To evaluate the proposed method an implementation was made and subsequently validated using the data set described in Section 11.3. The implementation was done in C++ as an extension of our AAM framework (see [213]) and executed on a 2.4 GHz Pentium PC. The stable period was set manually to the last 25 frames of each sequence, i.e. $S = 26$, except for patient 1 who had a delayed bolus passage, $S = 36$ was used. (This parameter can presumably easily be set automatically using the extremum of the temporal image derivative.) Due to the limited data set size, a leave-one-out cross validation was carried out producing ten test scenarios. The model was automatically initialised in the P-th frame using the approach described above.

Two performance benchmarks were calculated for each model landmark: *pt.pt.*, which measures the Euclidean distance between corresponding landmarks of the model and the ground truth; and *pt.crv.*, which measures the shortest distance to the ground truth curve. These are referred to as *landmark errors* in the following. When grouped into mean values for a shape these are referred to as *shape errors*. The initialisation strategy successfully located the heart in all patients, but number nine. The CAAM for this patient was subsequently hand-initialised and included on equal terms with the remaining nine.

Table 11.2 summarises the shape errors for all patients, while Table 11.3 expands on these numbers to show errors for each patient. Computation timings for the initialisation and the image search are given in Table 11.4.

Shape errors versus frame number are depicted in Figure 11.6 and 11.7 as mean values with error bars and as box plots, respectively. Notice the consistent increase of the shape error during the unstable period. Plotting shape errors versus patient

Table 11.2: Mean shape errors (shown in pixels).

	Mean	Std.dev.	Median	Min	Max
Pt.pt.	2.81	0.71	2.68	1.51	6.69
Pt.crv.	1.25	0.36	1.17	0.66	3.79

Table 11.3: Shape errors for each patient (shown in pixels).

Patient	1	2	3	4	5	6	7	8	9	10
					Pt.pt.					
Mean	2.50	2.81	3.45	3.23	2.68	2.53	2.62	3.26	2.75	2.26
Std.dev.	0.46	0.71	0.44	0.83	0.41	0.73	0.63	0.39	0.90	0.35
Median	2.45	2.74	3.45	2.99	2.62	2.40	2.52	3.28	2.59	2.21
Min	1.51	1.76	2.53	2.21	1.90	1.52	1.73	2.49	1.74	1.73
Max	3.50	4.56	4.39	5.13	3.72	5.34	5.14	4.06	6.69	3.72
					Pt.crv.					
Mean	1.29	1.07	1.62	1.31	1.03	1.18	1.35	1.21	1.35	1.10
Std.dev.	0.23	0.15	0.35	0.22	0.21	0.48	0.39	0.16	0.54	0.24
Median	1.27	1.03	1.68	1.26	0.97	1.05	1.27	1.18	1.21	1.06
Min	0.78	0.85	0.86	1.03	0.75	0.66	0.98	0.94	0.80	0.83
Max	1.82	1.50	2.20	1.83	1.69	2.98	3.26	1.67	3.79	2.21

Table 11.4: Timings (shown in seconds).

	Sequence time	Mean frame time
Initialisation	34.0	0.7
CAAM search	9.9	0.2
Total	43.9	0.9

number in Figure 11.8 shows approximately 14 apparent search failures out of 500 frames (14/500=2.8%). From Figure 11.6 it is seen that all of these occurred during the unstable period. As indicated by circles in Figure 11.8, seven of these were detected and flagged by Equation 11.10. For patients 1–3 the replaced mean estimator was close to the ground truth shape, contrary to patient 9.

To give a qualitative impression of the accuracy Figure 11.9 and 11.10 show registration results before, during and after the bolus passage for patient 10 and 8, respectively. Notice the dark region showing a severe perfusion deficit present in the anterior and inferior septal wall of patient 8 in Figure 11.10.

To complete the quantitative overview, Figure 11.11 shows the distribution of pt.crv. landmark errors in addition the their cumulative frequency. It is seen that approximately 80 percent of the landmark errors are below two pixels. Finally, Figure 11.12 illustrates the spatial distribution of the landmark errors by showing circles with radii proportional to the mean landmark error on the mean shape. Errors on RV dominate together with errors on the LV endocardial contour in proximity of the papillary muscles, i.e. near the anterior and inferior lateral wall, of the first and third slice.

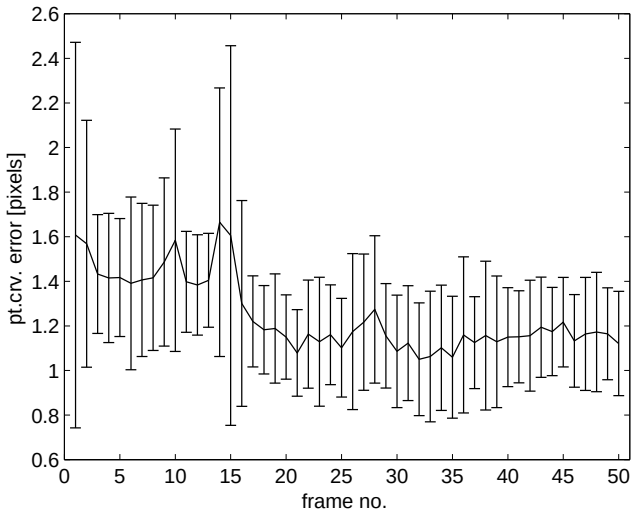


Figure 11.6: Pt.crv. shape errors for each frame. Error bars are one std. dev.

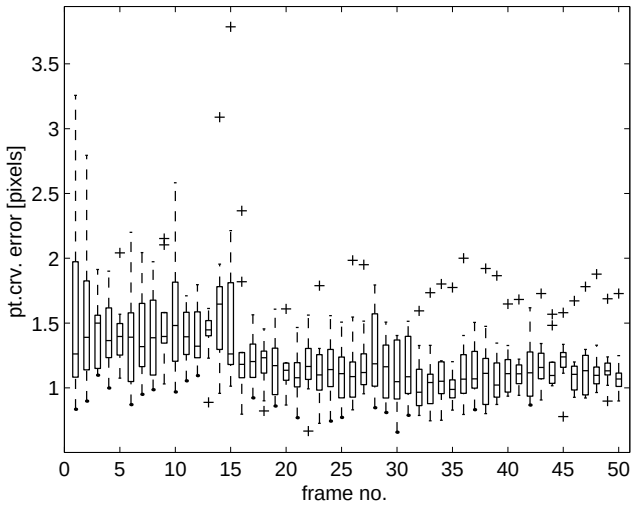


Figure 11.7: Box plots of the pt.crv. shape errors for each frame (IQR=1.5).

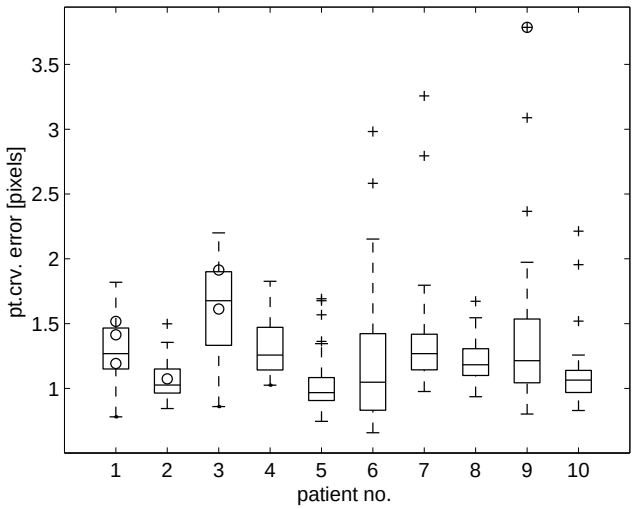


Figure 11.8: Box plots of the pt.crv. shape errors for each patient (IQR=1.5). Pose prior enforcements are shown using circles.

Figure 11.9: Registration results for patient 10 before, during and after bolus passage.

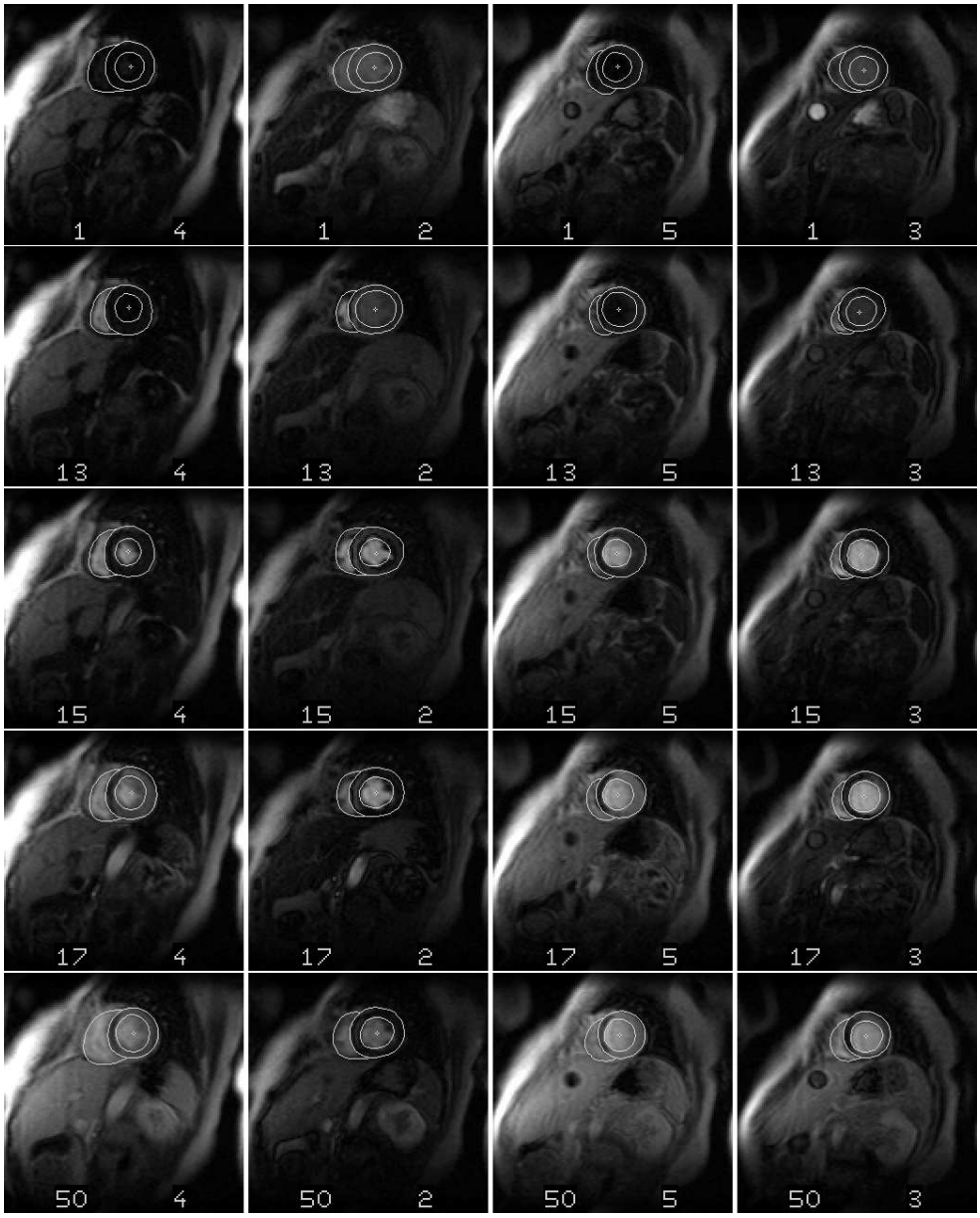


Figure 11.10: Registration results for patient 8 before, during and after bolus passage.

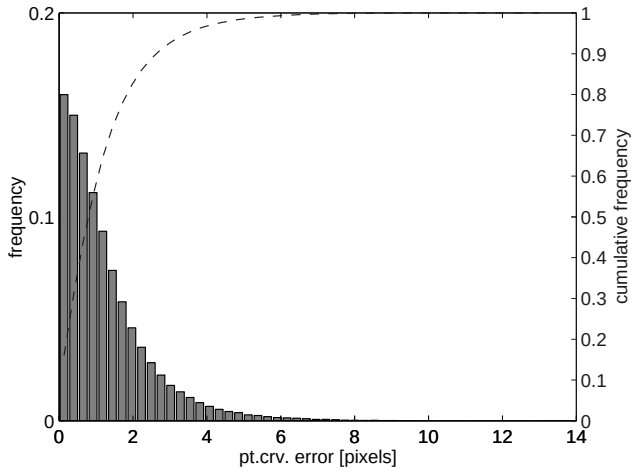


Figure 11.11: Distribution of pt.crv. landmark errors for all frames.

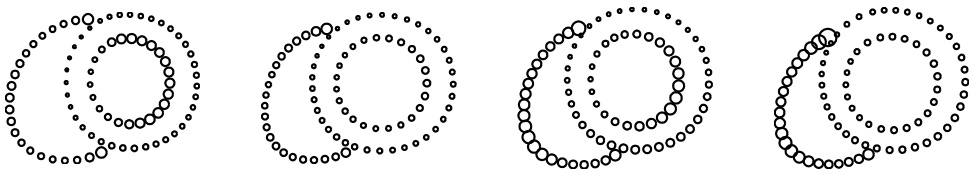


Figure 11.12: Mean pt.crv. landmark errors plotted as circles on the mean shape of each slice.

11.6 Discussion

Registration of perfusion MRI remains a challenge to medical image analysis. However, we believe that the presented method holds promise for the future, although conclusions based on the limited data material will remain fragile. This optimism is partly due to the ability to produce useable results on pathological cases – all patients were diagnosed with acute myocardial infarction – from MRI scans exhibiting an image quality below average compared to recent developments in MR cardiac imaging.

Registration results were produced without manual interaction in nine out of ten cases, due to the described initialisation method. This strategy failed for patient 9, for whom the heart was poorly centred in the slice field-of-view. Further, and more importantly, slices were shifted heavily towards the apex compared to the remaining nine patients, which obstructed a good fit of the LV-RV model in the initialisation process. We consider this patient an outlier and accept the observed limitations.

In general, less than three percent model fit failures were observed judged from Figure 11.8. All occurred prior to, or during the bolus passage. Half of these were not detected by Equation 11.10. We consider this performance convincing, although it indicates that registrations based on this limited training data should be reviewed by the operator, prior to making judgements upon automatically generated perfusion maps. The operator workload involved herein corresponds approximately to performing manual annotation of a single frame.

Figure 11.12 reveals that a substantial part of the landmark error stems from imperfections in the localisation of the right ventricle. LV registration errors are consequently smaller than the reported averages in the previous section. Although the inclusion of the RV adds specificity to the presented model, a hierarchical approach using a separate LV model initialised from the converged RV-LV model is straightforward and should be well worth to pursue.

Since only weak assumptions regarding sequence length and bolus passage position are present in the method, this is easily adapted to arbitrary-length perfusion sequences. The method only requires a coarse indication of the bolus passage position and a stable period of a reasonable number of frames after the bolus, the latter being a standard requirement in perfusion MRI. As stated earlier, the former should be possible to obtain automatically from the extremum of the temporal image derivative.

We believe that the presented method has a noteworthy set of merits compared to other approaches. i) It is based upon a well-understood and well-described framework; the Active Appearance Models, which also has a freely available implementation. ii) It has met the robustness requirements crucial when dealing with perfusion MRI. iii) The implementation showed that results could be produced within a reasonable time frame. iv) Imperfect ECG-triggering is handled by allowing the

model to perform complex deformations. v) Although current MR pulse sequences do not allow for multi-slice perfusion sequences without ECG-triggering, this can also be handled by the deformation properties of our method. Finally, vi) To let our method generalise to new data, very few assumptions concerning the data content have been made. Except for a few scalar parameters – which are dimensionless indices relating to the statistics of the actual data – all values are estimated from training data, rather than being explicitly coded into the method. We believe this to be a very fruitful approach, as the method easily adapts to new expert knowledge given by medical doctors. Knowledge, that typically already exists in the form of hand-annotated training data from previous studies.

However, the above virtues come at a price. Image annotation is not avoided altogether. A training set needs to be carefully crafted in a similar process to the one this method is aiming at eliminating. However, we have taken steps to alleviate this part significantly by introducing spline-based curve fitting and MDL shape modelling into the annotation process. Further, even if a method requires no training data, a validation set still needs to be established. In summary, we do not consider this issue a major Achilles heel. Secondly, this framework is built upon assumptions of correspondence. Consequently, papillary muscles, which are present in some cases and seemingly not in others, cannot be modelled and thus provides a limit on the level of detail obtainable by such methods.

Additionally, compared to other approaches given in Table 11.1 our method has the advantage of modelling all slices simultaneously. Furthermore, our computation times are well competitive, keeping in mind that they apply to modelling of four slices instead of one. Some of the approaches only correct for translational motion, disregarding rotation and deformable shape changes. Our deformable modelling approach gives an accurate registration – a dense point correspondence through the whole sequence. Further, we have given both qualitative and quantitative validation of our method involving comparison with manually defined (ground truth) myocardial boundaries. This is a clear advantage to most of the other approaches, which rely on relative comparison of parameters derived from the raw and the registered data.

Compared to our previous work in [224] several extensions have been introduced. A larger data set containing multiple slices from patients, compared to single slices from controls, has been obtained. Image annotation and shape models are substantially improved using spline interpolation and MDL optimised shape models. Slice coherence is modelled compared to single slice-oriented models. Models are utilising the full dynamic range of the MR scanner compared to the truncated 8-bit representation available previously. Automated model initialisation is introduced to avoid the need for any manual interaction. Cardiac interfaces to the upper abdomen and the lung are now included into the texture model to improve specificity.

The initialisation part aside, all presented results here are produced reasonably rapidly, albeit not as rapid as in the previous work. This has several reasons. Models are substantially larger as they encompass several slices and include the cardiac

interfaces as well. Images have twice the bit depth, which means more memory transfer, less CPU cache consistency, et cetera. Finally, there has been put far less emphasis on code optimisation compared to the previous work. We anticipate that substantial improvements can be achieved on this level alone.

We intend to expand the current work in the following areas. While the need for clustered texture models is obvious, the presented results still leave room for improvement. This is also true for the brute force model selection used, which makes search time (approximately) linear in the number of classes. Model initialisation could be improved by sampling the pose distribution of the training set and by using a multi-resolution implementation, which both will decrease the initialisation time substantially. Further, the claims regarding automatic bolus detection should be proven correct. But most importantly, a clinical validation on a much larger training set should be carried out to demonstrate the true capabilities of the presented statistical approach and its alleged adaptive properties. In addition, it should be explored to what extent LV sub models – as mentioned earlier – will improve the accuracy of myocardium registration. Performance evaluation should be extended to include the clinical end goal; the various generated perfusion maps. Consequently, estimated perfusion parameters should be compared as obtained from automatic versus hand-registered sequences.

11.7 Conclusion

We have described a novel, data-driven method for registration and motion-compensation of multi-slice cardiac perfusion MRI. Preliminary validation of the method on ten patients with acute myocardial infarction showed promising results. In nine cases results were obtained without manual interaction. Despite evident perfusion deficits and varying image quality in the limited training set our method yielded a mean point to curve distance of 1.25 ± 0.36 pixels for the left and right ventricle combined. We conclude that this learning-based method holds great promise for the automation of cardiac perfusion investigations, due to its accuracy, robustness and generalisation ability.

11.A Estimation of $\mathbf{\Sigma}$ and σ

The dispersion matrix of the pose parameters, $\mathbf{\Sigma}$, and the standard deviations of the shape parameters, σ, can be estimated from the stable period. However, in order to obtain more reliable estimates we propose to use the training sequences, since many additional samples are available. To avoid confusion of inter- and intra-sequence variability, we filter out sequence-specific information (such as mean shape size) prior to the estimation. Now, let R denote the number of sequences, each containing P frames. Further, let $\odot$ denote the Hadamard, i.e. element-wise, product. Then the estimation of $\mathbf{\Sigma}$ and σ can be specified as shown in Algorithm 11.2.

Algorithm 11.2 Estimation of pose and shape variation

1: **for** $i = 1$ to R (for each training sequence) **do**
2: $\quad \mathbf{p}_t \leftarrow \mathbf{p}_t \odot (\frac{1}{\overline{s}}[\, 1\ 1\ 1\ \overline{s}\,]^\mathsf{T})\ \forall\ t$ — Normalise pose w.r.t. mean size
3: $\quad \overline{\mathbf{p}}_i = \frac{1}{P}\sum_{t=1}^{P} \mathbf{p}_t$ — Estimate pose seq. mean
4: $\quad \mathbf{\Sigma}_i = \frac{1}{P-1}\sum_{t=1}^{P}(\mathbf{p}_t - \overline{\mathbf{p}}_i)(\mathbf{p}_t - \overline{\mathbf{p}}_i)^\mathsf{T}$ — Estimate pose dispersion matrix
5: $\quad \overline{\mathbf{b}}_{s,i} = \frac{1}{P}\sum_{t=1}^{P} \mathbf{b}_{s,t}$ — Estimate shape mean
6: $\quad \sigma_i = \sqrt{\frac{1}{P-1}\sum_{t=1}^{P}(\mathbf{b}_{s,t} - \overline{\mathbf{b}}_{s,i}) \odot (\mathbf{b}_{s,t} - \overline{\mathbf{b}}_{s,i})}$ — Estimate seq. shape std. dev.
7: **end for**
8: $\mathbf{\Sigma} = \frac{1}{R}\sum_{i=1}^{R} \mathbf{\Sigma}_i$ — Calculate pooled estimate for all sequences
9: $\sigma = \frac{1}{R}\sum_{i=1}^{R} \sigma_i$ — Calculate pooled estimate for all sequences

Further, to maintain simplicity, normalisation with respect to rotation has not been included in the above algorithm. However, if sequences differ substantially in orientation, this should be taken into account when estimating t_x and t_y. Finally, during CAAM search, appropriate normalisation and de-normalisation using the stable period mean size (and orientation) of the current sequence must be carried out.

CHAPTER 12

Rapid and Unsupervised Correction of Respiratory-induced Motion in 4D Cardiac Cine MRI

Mikkel B. Stegmann and Henrik B. W. Larsson

Abstract

This paper describes and evaluates a rapid and unsupervised method for correction of respiratory-induced motion in dynamic three-dimensional cardiac magnetic resonance images. As such, the method provides a purely image-based alternative to e.g. respiratory gating. Statistical models of shape and appearance are employed to localise the left and right ventricle in a short-axis slice between the base and the apex. The segmented region is subsequently propagated through all slices in the apex-basal direction using a non-rigid registration process, thus providing an estimate of the inter-slice translation due to variation in inspiration depth for each breath-hold. Qualitatively and quantitatively validation carried out on 12 subjects appear very promising. We found that slice displacements were significantly reduced in the corrected images. Joint estimation and correction took less than 10 seconds per subject.

12.1 Introduction

Due to sizable research efforts in improving the acquisition of magnetic resonance images (MRI), investigations of dynamic phenomena such as the beating heart are now possible. Still, in order to obtain acceptable signal-to-noise ratios different creative acquisition schemes must be employed. In the case of dynamic cardiac imaging, information is typically accumulated over many heartbeats to form an image sequence covering the complete heart cycle. Temporal correspondence is ensured by electrocardiogram (ECG) triggering, which starts acquisition relative to the R-peak.[1] To succeed, this accumulation approach has two major requirements: i) that no global movement of the heart occurs during acquisition and, ii) that the heart rate remains constant during acquisition. While the latter is typically less of a problem for patients without arrhythmia, the former prohibits breathing during acquisition due to the substantial displacement of the heart during the respiratory cycle, which would result in severe image corruption as image information is integrated from data not in correspondence.

Prohibiting breathing puts obvious restrictions on the length of the acquisition period. Consequently, full spatial coverage of the heart is usually obtained by using series of breath-holds; each resulting in a single 2D time-series covering a particular spatial region of the heart. Each such time-series will be referred to as *phases* of a particular *slice*. Figure 12.1 shows a four-dimensional cardiac image composed of a stack of such slices. Here, the phase corresponding to the end-diastole is shown together with the geometry of the left ventricle (LV), which is given manually by a medical expert. The slice plane (xy-plane) is showing the abdomen in the lower part, while the central part contains the heart and lung. The two minor planes (xz-plane and yz-plane) show image data across slices. Notice the anisotropy of the voxels. Refer to Figure 12.2 for a schematic of the described acquisition process used to obtain Figure 12.1.

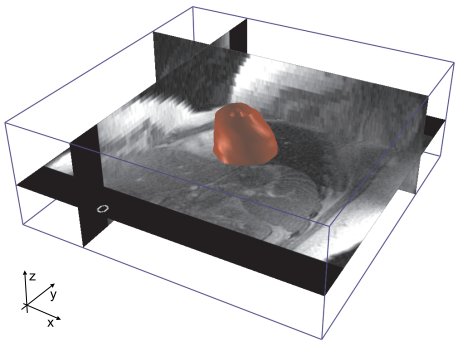


Figure 12.1: Four-dimensional cardiac cine image shown at end-diastole. Ground truth endocardial and epicardial surfaces are shown as shaded meshes.

[1] The most prominent peak in an ECG.

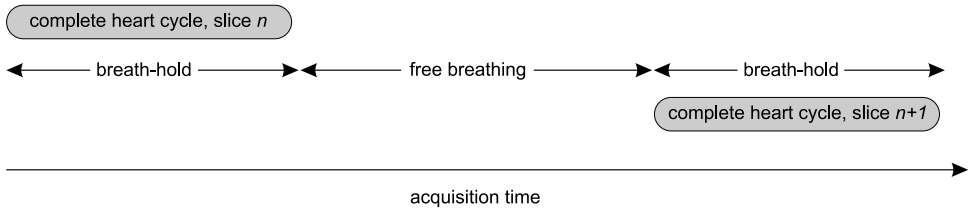


Figure 12.2: Schematic of dynamic, three-dimensional acquisition of cardiac MRI.

Besides a prolonged acquisition time, the obtained 3D volumes from this process will typically not show consistent cardiac anatomy, irrespective of the outer fixation of the subject in the MR scanner, due to the variation in respiration depth between breath-holds. This variation will thus cause the diaphragm (the upper abdomen) and the heart to reside in different areas of the image plane. This behaviour is shown in Figure 12.3 and acknowledged by many, e.g. Mitchell et al. [161] states:

> *... However, the MR data sets are acquired over several heartbeats as sequences of 2-D slices, not in a volumetric fashion. Individual slices are frequently acquired during separate breath-holds. Differences in inspiration level and thus heart position lead to variable shifts in the LV position between slices.*

Or, in the formulation of Moses and Axel [166]:

> *Inter-slice movement posed a significant problem in obtaining accurate descriptions of the septa. This is because each slice is acquired during a single breath-hold and thus, with different depths of inspiration, the heart occupies a different position in space, resulting in relative movement between the slices. To minimize this error we restricted our analysis to the long axis series, as the positional inconsistencies were less of a problem in this plane. However, this resulted in undersampling of the septum.*

Similar views are expressed in [86] and [32].

This work focuses on a purely image-based solution to the problem of correcting for these random slice displacements. Ultimately, such a process should seek to recover the true geometry of the heart. However, as we will touch upon later, this is an involved process and some approximations must be accepted.

Alternatives to an image-based method includes various ingenious solutions based on hardware such as the *stretch belt* (also called respiratory ”bellows”), which is attached to the subject providing information regarding the inspirational state to the MR scanner. This signal can synchronise the acquisition to the respiratory cycle – typically near the end-expiratory or end-inspiratory phase. Unfortunately,

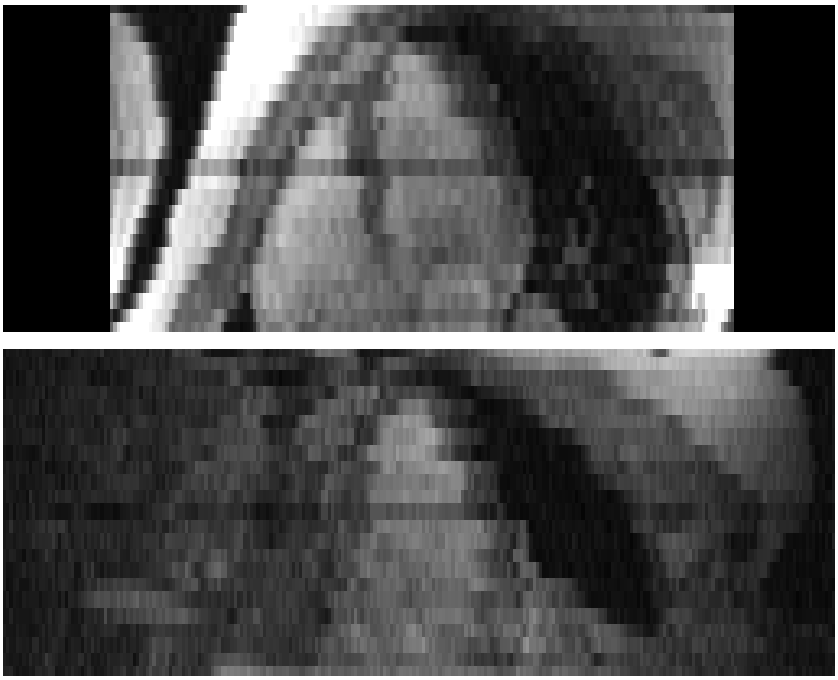


Figure 12.3: Distortion of cardiac anatomy due to respiration, xz-plane (top) and yz-plane (bottom). Notice the consistent lung border to the right, which is not affected by respiration.

such *respiratory gating* prolongs the imaging time significantly as only a small time window in each respiration cycle can be used for acquisition. An elegant substitute for the stretch belt is the *navigator echo* technique [90], where the respiratory state is estimated by a quick prepulse MR sequence that measures the position of the diaphragm before collecting image data.

A fully unsupervised image-based method holds several advantages to the above. It does not consume any imaging time and it does not require any manual interactions, contrary to the navigator echo and the stretch belt approaches. Most importantly, it does not require per-subject calibration of any intrinsic parameters, such as gating thresholds, trigger acceptance window, and diaphragm position, see e.g. [86]. Finally, it can be carried out retrospectively. However, a purely image-based solution also possesses some limitations, which will be discussed in the later sections.

12.2 Data Material

The cardiac MRI data used for this study consists of two independent sets. Set A consists of 14 spatially corresponding, short-axis, end-diastolic, 2D slices acquired

from 14 normals. All slices were positioned centrally between the base and the apex. Set B consists of 12 short-axis, dynamic volumes obtained from 12 obese normals (having a body mass index larger than 33 $\mathrm{kg/m^2}$). Both sets were acquired using a whole-body MR unit (Siemens Impact) operating at 1.0 Tesla. The used pulse sequence was an ECG-triggered, breath-hold, fast low angle shot (FLASH) cinematographic pulse sequence. Matrix=256×256, field of view=263×350 mm, slice thickness=10 mm (set A) and 6 mm (set B), phase time=55 ms. The volumes of set B had an average size of $(x, y, z, phases)$ = 256×256×23×12 voxels.

The endocardial and epicardial contours of the left ventricle in both sets were annotated manually by placing landmarks along each contour, see Figure 12.4. To fix rotation around the left ventricular (LV) long-axis, the right ventricle (RV) was also annotated in set A. Prior to further processing, subject 11 in set A was removed due to a scanner calibration error. In set B, only the end-diastole (ED) and the end-systole (ES) were annotated. All annotations were carried out in 2D in the xy slice plane.

Figure 12.4: Example annotation of the left ventricle in set A using 66 landmarks.

12.3 Methods

12.3.1 Problem Statement

Let $\mathcal{V}$ denote an image defined in four dimensions. In our application is it convenient to regard $\mathcal{V}$ as composed by S two-dimensional image slices acquired at P time phases. Let each slice image have the width and height of W and H, respectively. Now let $x = 1 \ldots W$, $y = 1 \ldots H$, $z = 1 \ldots S$, and $t = 1 \ldots P$. Further, let a position in the image be denoted by the vector $\mathbf{x} = [\ x\ y\ z\ t\ ]^\mathsf{T}$. Each $\mathcal{V}$ covers

exactly one heart cycle in temporal direction.[2] By convention, $t = 0$ is identical to the end-diastole.

At this point we will model effects from respiration as purely in-plane translations. Thus, the problem of correcting for slice displacements can be stated as an estimation of the vector-valued warping function, $\mathbf{f}(\cdot)$, in

$$\mathcal{V}_{corrected}(\mathbf{x}) = \mathcal{V}(\mathbf{f}(\mathbf{x})), \tag{12.1}$$

which takes on the following particular simple form

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} x - f_x(z) \\ y - f_y(z) \\ z \\ t \end{bmatrix}. \tag{12.2}$$

Ultimately, we would like to find a cost function, $d(\cdot)$, which has a well-defined global minimum where the true heart geometry is recovered in $\mathcal{V}$ corrected by applying $\mathbf{f}$. Provided such a function can be found – and time consumption is not an issue – the sought solution can be found by a standard general-purpose optimiser[3] by

$$\arg\min_{\mathbf{f}} \; d(\mathbf{f}, \mathcal{V}). \tag{12.3}$$

However, defining such a cost function is not a simple task. First, it should be made clear that it cannot reliably be based on 4D images similar to those in set B, since none of these would contain truly, three-dimensional heart geometry. Or, slightly rephrased; to construct $d(\cdot)$ so that it has a "correct" minimum; actual heart geometry must be known, which could be achieved by engineering detailed anatomical priors of the heart. One solution to this could be models of shape and appearance (see e.g. [54, 161, 213]) based on histological cuts and MRI. Unfortunately, such a database is currently not available. It should be noticed that recent advances in real-time MR imaging [32] hold promise for high definition 3D cardiac images acquired in one breath-hold. However, such scans are currently limited in terms of spatial resolution and signal-to-noise ratio, and were not available for the subjects in this study.

A less involved – and practically much more feasible – alternative is to explicitly state a prior, which only approximates the heart geometry. Solutions to Equation 12.3 will consequently also only approximate the actual displacements. Let us introduce the *apparent slice displacement*, $\mathbf{h}(\cdot)$, composed of the respiratory displacement, $\mathbf{f}(\cdot)$, and the contribution from changes in LV anatomy along the LV long-axis,

[2] Or more precisely, exact up to the accuracy given by the fixed temporal sampling, here 55 ms.

[3] E.g. simulated annealing in a Metropolis-Hastings sampling scheme.

$\mathbf{g}(\cdot)$,

$$\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x}). \tag{12.4}$$

We chose to define the $\mathbf{h}(\cdot)$, as the displacement of the centre-of-gravity (COG) of the endocardial contour. Further, we adopt the definition of LV long-axis given by Lelieveldt et al. [145, 148], which is the line fitted through all LV endocardial COGs. Hence, if the short-axis scan planning is optimal, image slices will be orthogonal to the long-axis and consequently $\mathbf{g}(\mathbf{x})$ should be zero on average.

Thus, we propose to disregard $\mathbf{g}(\mathbf{x})$ and accept the approximation $\mathbf{f}(\mathbf{x}) \approx \mathbf{h}(\mathbf{x})$. One consequence of this is, that any consistent skewness across slices – e.g. stemming from poor short-axis planning – will be eliminated. However, we will later describe how such rectification can be avoided altogether in a post-processing step.

In practice, the apparent slice displacement will be estimated from the translation between slices that minimise changes in image appearance of the LV from one slice to another, $(\phi_x(z), \phi_y(z))$. With a minor abuse of notation our cost function at time phase t becomes

$$d_t(\mathbf{f}, \mathcal{V}) = \sum_{z=2}^{S} \rho\left(\mathcal{V}(\mathbf{f}(\mathbf{\Omega}_{t,z})), \mathcal{V}(\mathbf{\Omega}_{t,z-1})\right), \tag{12.5}$$

where $\mathbf{\Omega}_{t,z}$ denotes the LV heart region in the slice z at time phase t and $\rho(\cdot,\cdot)$ is a similarity metric between two sets of voxel values.

Thus, two problems remain. As the goal of this work is an unsupervised method, $\{\mathbf{\Omega}_{t,z}\}_{z=1}^{S}$ must be determined automatically. Secondly, correspondences between differently sized LV regions must be established by ρ.

12.3.2 An Example Solution

In the remaining part of this report we will investigate one particular simple and computationally efficient solution to the problem stated above.

As a baseline solution we propose to localise the left and right ventricle (LV, RV) of the heart in a single slice (centrally between the base and the apex), and propagate this solution from slice to slice in the apex direction and subsequently in the basal direction, to establish the required heart regions; $\{\mathbf{\Omega}_{t,z}\}_{z=1}^{S}$. This sequential propagation should be carried out using ρ as similarity measure. Thus, the translational components between $\mathbf{\Omega}_{t,z}$ and $\mathbf{\Omega}_{t,z-1}$ for $z = 2 \ldots S$ will implicitly provide the solution to Equation 12.3 since only neighbouring slices are considered in Equation 12.5. It should be emphasized that this solution is only valid to the extent that $\{\mathbf{\Omega}_{t,z}\}_{z=1}^{S}$ denotes the actual heart regions. Any localisation errors, will result in erroneous estimates of slice displacements.

To localise the LV and RV, we employ the learning-based Active Appearance Models (AAMs) [52, 54] framework for image interpretation. In this application a cardiac AAM is trained on the end-diastolic images in set A. Using a multi-resolution initialisation method described in [205], enables it to reliably locate the heart in a single

Figure 12.5: Unsupervised localisation of end-diastolic LV and RV contours in subject 4.

short-axis image within a fraction of a second. This method has previously been validated and described in details in [213, 226].

As the position of the sought after central slice between the base and the apex are not fixed in general, the AAM searches the central 50% of all end-diastolic slices in each 4D cardiac MRI to localise the heart. The slice result providing the best[4] fit to the model is chosen as the initial localisation. Refer to Figure 12.5 for an example. Due to the substantial changes in LV and – in particular – RV appearance from base to apex, only a reference region of interest around the LV is extracted. In the current work, we simply use an inflated axes-aligned bounding box determined from the epicardial[5] contour. Such inflation is required to establish a neighbourhood context outside the epicardial contour.

Subsequently, this region of interest (ROI) can be propagated through every frame. This is carried out in the apex direction first, and next in the basal direction. Propagation is carried out by a search in the space of Euclidean similarity transformations, i.e. translation, scaling and rotation of the ROI. Thus, this search establishes a dense correspondence between the two ROIs by maximising their similarity. In practice, the search is carried out efficiently by iteratively applying parameter updates based on a pre-computed gradient matrix. See [54, 65] for the details. Figure 12.6 shows the results of such a propagation. Displacements between ROIs are measured as the displacement of their corresponding centroids.

Since the true slice displacements are constant *per se* for all time phases – see e.g. Equation 12.2 – the above procedure can be carried out independently for $t = 1 \ldots P$. Displacement results estimated independently from each cardiac phase

[4] In a least squares sense, $\rho(model, image) = ||model - image||_2^2$.

[5] The outer contour the left ventricle.

Figure 12.6: Slice displacement estimation at end-diastole in subject 4 by propagation of the reference ROI shown in slice 8. Images are shown row-wise from base to apex.

Algorithm 12.1 Estimation of slice-to-slice displacements

Require: $\mathcal{V}$ and an end-systolic, shot-axis, AAM
1: Let $\mathcal{V}(z,t)$ denote the 2D image at slice position z and time phase t
2: **for** $z \in$ the central 50% of all slices **do**
3: Perform a cardiac AAM search to localise the heart region in $\mathcal{V}(z,0)$
4: Store shape result, $\mathbf{s}_z$, and model-to-image fit, ϵ_z
5: **end for**
6: Determine the slice with the best search result; $Z = \arg\min_z \epsilon_z$
7: Extract the reference LV region of interest $\mathbf{\Omega}_{ref}$ from s_Z
8: Initialise reference regions; $\mathbf{\Omega}_{t,Z} = \mathbf{\Omega}_{ref}\ \forall\, t$
9: **for** $z = Z+1$ to S **do**
10: **for** $t = 1$ to P **do**
11: Estimate $\mathbf{\Omega}_{t,z}$ by matching $\mathcal{V}(\mathbf{\Omega}_{t,z-1})$ to $\mathcal{V}(z,t)$
12: **end for**
13: Fuse information from $\{(\mathbf{\Omega}_{t,z}, \mathbf{\Omega}_{t,z-1})\}_{t=1}^{P}$ to form inter-slice displacement $\hat{\phi}_x(z), \hat{\phi}_y(z)$
14: **end for**
15: **for** $z = Z-1$ down to 1 **do**
16: **for** $t = 1$ to P **do**
17: Estimate $\mathbf{\Omega}_{t,z}$ by matching $\mathcal{V}(\mathbf{\Omega}_{t,z+1})$ to $\mathcal{V}(z,t)$
18: **end for**
19: Fuse information from $\{(\mathbf{\Omega}_{t,z+1}, \mathbf{\Omega}_{t,z})\}_{t=1}^{P}$ to form inter-slice displacement $\hat{\phi}_x(z+1), \hat{\phi}_y(z+1)$
20: **end for**
21: $\hat{\phi}_x(1) = 0$
22: $\hat{\phi}_y(1) = 0$

can thus be fused to provide possibly more robust estimates. Since the rectangular model has no notion of the LV geometry and the contractile pattern of the heart induces little variation of the epicardial contour, the ROI determined at end-diastole is used for all time phases in this setup.

The complete procedure is expressed as pseudo-code in Algorithm 12.1. The actual fusion of the displacements in line 13 and 19 should optimally depend on the distribution of the error in the estimates. In the current work we have chosen to evaluate three strategies, i) no fusion: use the displacement from a single fixed time phase (denoted by '*t*' below), ii) use the result from the time phase providing the best model-to-image fit (denoted '*of*' below), and iii) use the mean displacement of all time phases (denoted '*me*' below).

From the notation of Algorithm 12.1, and provided that it is acceptable to ignore any skewness, the estimate of the correcting warping function stated in Equation 12.2 can then be stated as

$$\hat{\mathbf{f}}(\mathbf{x}) = \begin{bmatrix} x - \hat{f}_x(z) \\ y - \hat{f}_y(z) \\ z \\ t \end{bmatrix} = \begin{bmatrix} x - \sum_{i=1}^{z} \hat{\phi}_x(i) \\ y - \sum_{i=1}^{z} \hat{\phi}_y(i) \\ z \\ t \end{bmatrix}. \tag{12.6}$$

12.4 Experimental Results

All twelve dynamic volumes in set B have been corrected using the method described above. The only auxiliary data required was the Active Appearance Model built on data set A.

To assess the method quantitatively, end-diastolic LV contours have been annotated manually in all images in set B. Due to the definition of the apparent slice displacement above, a direct measure the efficacy of the method is given by calculating the centre of gravity (COG) displacement of the endocardial contour from slice to slice, before and after correction. We stress that this measure states nothing about the quality of the correction outside the LV region.

Figure 12.7 shows the distribution of these displacements before and after correction for all twelve subjects. In this figure, displacements have been estimated using the end-diastolic phase only, $t = 0$. Aside from a significant reduction in the mean displacement, the less tail-heavy distribution should be appreciated. The corresponding mean and standard deviation of the displacements calculated from original and corrected images were in this case 3.44$\pm$3.09 mm and 1.60$\pm$1.51 mm, respectively. The maximum displacement before and after correction was 22.7 mm and 8.9 mm, respectively.

Figure 12.8 and Table 12.1 show the corresponding subject means before and after correction. Notice the large variation of the mean displacement between subjects before correction. The corresponding statistics for all three fusion strategies described in Section 12.3.2 are shown in Table 12.2 and as box plots in Figure 12.9. Since the number of slices is varying from subject to subject (due to the differences in heart size and the fixed slice thickness), the mean values in Table 12.1 are not comparable to the mean values for method '0' in Table 12.2. The non-overlapping notches around the medians in Figure 12.9 indicate a significant improvement for all correction methods compared to the original images. However, a clear dependence between time phase and accuracy is observed, yielding a minimum at $t = 0$ as seen from Table 12.2. Using this single phase also outperformed the two fusion methods.

Figure 12.10 shows a qualitative comparison between the original and corrected image of subject 4. To aid the subjective assessment of the solution, ground truth endo- and epicardial geometry is shown as intersections with the volume planes, and further in two different projections. (This geometry is for validation purposes only and the presented algorithm was obviously blinded to this ground truth solution.) Notice the obvious reconstruction of cardiac and upper abdominal structures in the right column, while the respiratory-displacements have been "transferred" to the structures not affected by respiration, e.g. the rightmost part of images in row 1 and 2. The topmost image in the right column also reveals the rectification part of the solution, as the low frequency trend seen in the interface between the image data and the two encapsulating vertical black borders. Presumably a scanner calibration error is revealed in this subject as a dark central slice. Finally, the two projections

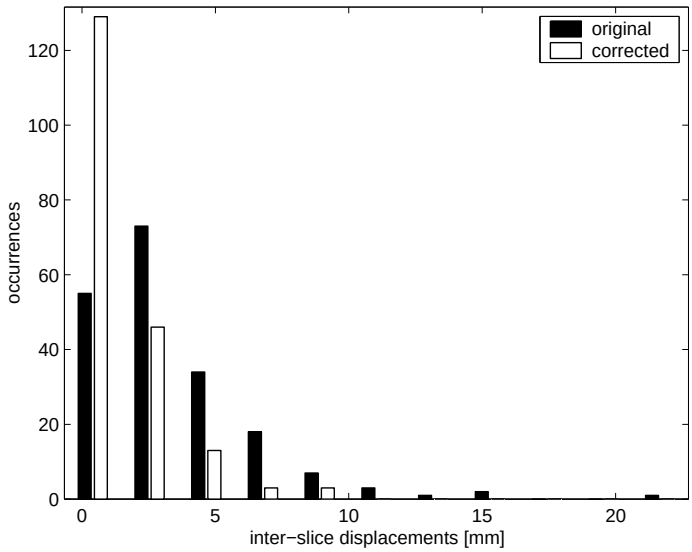


Figure 12.7: Histogram of slice displacements for all subjects before and after correction using method '0'.

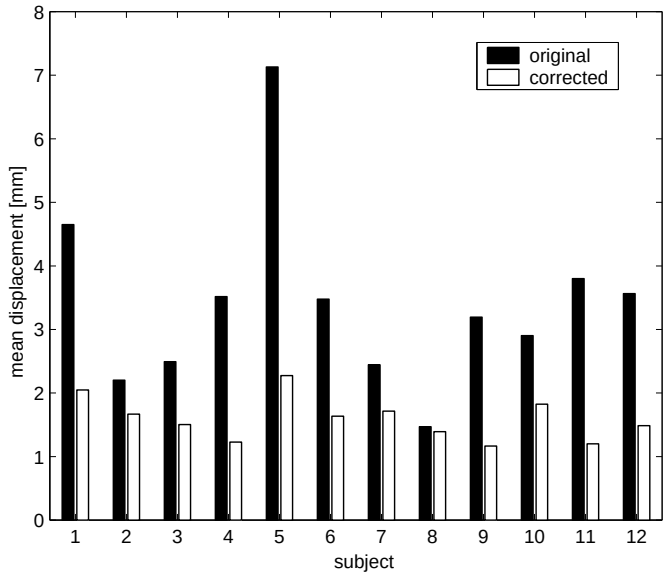


Figure 12.8: Mean slice displacements for each subject before and after correction using method '0'.

Table 12.1: Mean slice displacements for each subject before and after correction using method '0'.

Subject	Original [mm]	Corrected [mm]
1	4.65	2.05
2	2.20	1.67
3	2.49	1.50
4	3.52	1.23
5	7.13	2.27
6	3.48	1.64
7	2.44	1.72
8	1.47	1.39
9	3.19	1.17
10	2.90	1.83
11	3.80	1.20
12	3.56	1.49
Mean±std.dev.	3.40 ± 1.44	1.60 ± 0.34

Table 12.2: Mean and standard deviations for all slice displacements in all subjects. Shown for the original images and for each correction method.

Method	Displacement [mm]
none	3.44±3.09
0	1.60±1.51
1	1.71±1.71
2	2.05±2.24
3	2.03±1.99
4	2.13±2.02
5	2.06±1.84
6	1.97±1.86
7	1.80±1.71
8	1.80±1.66
9	1.79±1.64
10	1.70±1.63
of	1.67±1.72
me	1.68±1.65

of the geometry clearly shows that respiratory artefacts are most conspicuous along the y-axis. This is often the case due to the typical orientation of a short-axis image, see e.g. Figure 12.4. The subjective impact of our method is most pronounced when the heart-cycle is played in a cine-loop comparing original images to corrected ones. Movies of such can be delivered electronically by request to the first author.[6]

One application of having proper three-dimensional cardiac volumes that approximate the true tissue geometry is the possibility of building more accurate Point Distribution Models (PDMs) [68] in 3D for interpretation of unseen images, see e.g. [161, 162]. Resulting interpretation results could be used for inference of im-

[6]Or, by visiting http://www.imm.dtu.dk/∼aam/.

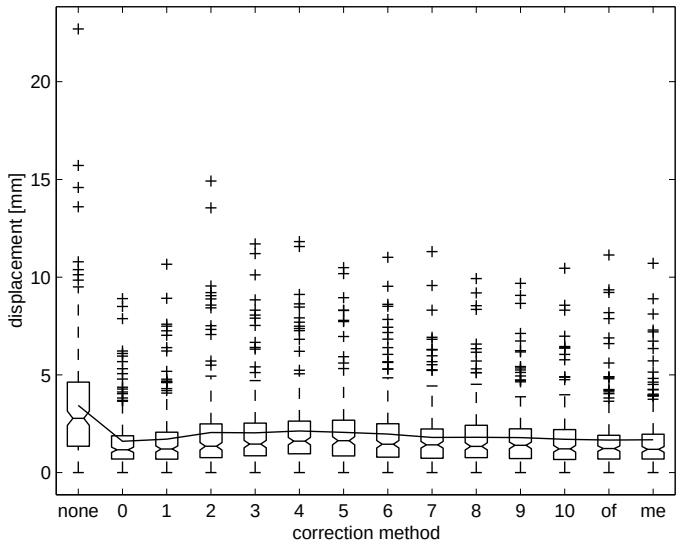


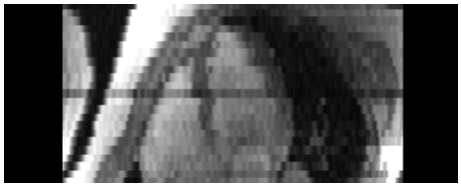
Figure 12.9: Boxplots of slice displacements for each correction method, IQR=1.5. The line through all methods shows the corresponding mean values.

portant functional indices such as the ejection fraction (EF), myocardium mass, et cetera. To assess the impact of our correction method three PDMs were built on original, corrected[7] and rectified annotations of the endocardial contours during end-diastole and end-systole and the epicardial contour during end-diastole. Rectified annotations were obtained by translating slice annotations to have a zero displacements by aligning the endocardial COGs. The relative variances of the resulting PDMs were 100%, 82%, and 79%, respectively. Hence, a significant increase in compactness was obtained for the automatically corrected shapes, comparing well to the ground truth rectified shapes.

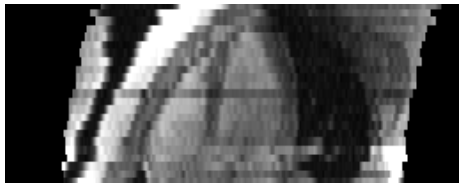
Our method was implemented in C++ and augmented by an Active Appearance Model framework described in [213]. In method '0', joint estimation and correction took less than 10 seconds per subject, on a 2.4 GHz Pentium 4 PC.

The used AAM framework and the data in Set A are both publicly available [207, 213], which should substantially ease a potential implementation of this method.

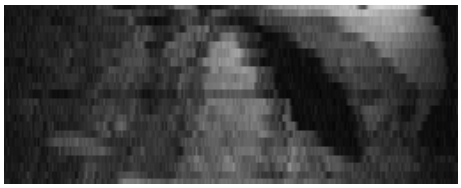
[7] Using method '0'.



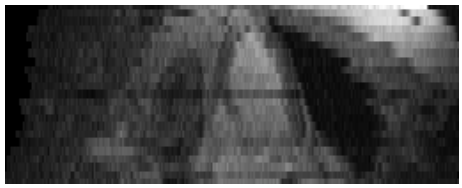
Original xz-plane.



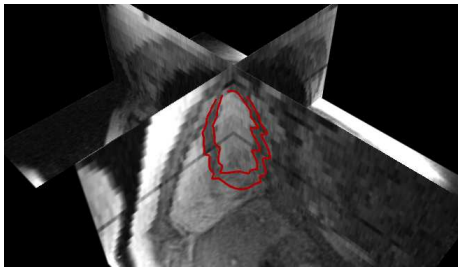
Corrected xz-plane.



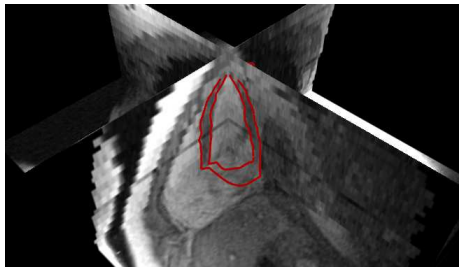
Original yz-plane.



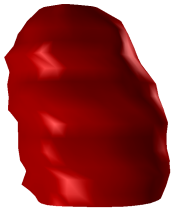
Corrected yz-plane.



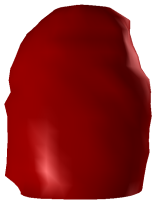
Original image with geometry intersections.



Corrected image with geometry intersections.

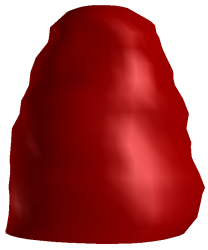


Original geometry in xz-projection.



Corrected geometry in xz-projection.

Original geometry in yz-projection.



Corrected geometry in yz-projection.

Figure 12.10: Qualitative comparison between the original and corrected 4D images for subject 4, here shown in the end-diastole. Rows 3–5 show the corresponding ground truth LV geometry with and without correction. Cutting planes in xz and yz were selected as being incident with the centre of gravity of the geometry.

12.5 Discussion

The merits of this work should be perceived as three-fold; a discussion of the problem at hand, a general description of the required parts of a solution, and finally an example solution designed to work rapidly and unsupervised. This section will discuss the details of the presented results as well as the potential weaknesses of the described method.

Looking at Figure 12.8 a considerable inter-subject variation is observed. This may be explained by variation in the motivation and cooperation of each subject, in addition to variation in how well each subject has been instructed in performing the breath-hold.

The results in Figure 12.9 may also surprise the naked eye. We hypothesise, that the increase in geometrical complexity of the myocardium towards the end-systole holds the key to this behaviour. As the muscle contracts the slice-coherence decreases and thereby leads the propagation process astray. Conversely, best performance is observed during end-diastole when the heart is maximally expanded, thus the myocardium shows the simplest geometrical appearance over the set of slices.

In the formulation of this method, several assumptions have been made. First, all respiratory effects are modelled as pure in-plane translations. While the in-plane component is the largest, through-plane effects exist. These are ignored in the current work.

Secondly, correcting apparent in-plane translations, which potentially can be composed of changes in the heart geometry yields – as mentioned previously – solutions that rectifies the myocardium. However, this potential rectification is minimised *per se* if the short-axis scan planning is accurate. Unfortunately, this is not always the case. The upper left image in Figure 12.10 clearly shows that the true long-axis of the left ventricle is not vertical as it – by definition – should be. One approach to suppress this rectification in poorly planned short-axis images would be to post-process the correction by a fit to the 3D orthogonal regression line estimated from the slice displacements. This might require an outlier rejection process as slice displacement estimates calculated outside the LV are notoriously error-prone in this LV-centred method. The fact that a line in 3D is given by only two observations, makes Random Sample Consensus (RANSAC) [94] an obvious candidate for robust line fitting. Alternatively, iterated re-weighted least squares (IRLS) [119] would also be well-suited as the fraction of potentially gross measurement errors is low, i.e. only pertinent to the slices outside the LV.

Finally, it should be stressed that the degree of adaptation employed by the propagation process is inherently prone to accumulation of errors. This fact alone, leads to a heteroscedastic noise on the displacement estimates having a parabolic shape with minimum in the slice where the reference ROI is located. This could

be exploited in e.g. the long-axis estimation procedure sketched in the previous paragraph.

The issues raised above may lead to the impression that the positive effects of the presented method are at its best only moderate and even may lead to new artefacts. This is not the case. Although error accumulation does influence the results, these still show a highly significant improvement over the original images. As such, we anticipate that using image-based correction similar to the presented yields much welcomed improvements on several fronts.

Cardiac image annotation is often carried out in a process where the medical expert integrates knowledge about the myocardial contours by flipping between the current, previous and next slice to resolve any ambiguities. Bringing slices into correspondence should thus improve both annotation accuracy, reproducibility and repeatability. Further, it should increase the productivity.

Removing the random effects of slice displacements will also make cardiac PDMs more compact and thereby more specific to the problem of analysing cardiac MRI. According to Occam's razor, we should expect these models to generalise better and consequently increase the segmentation/registration accuracy on unseen examples. The effect is hypothesized by Mitchell et al. [161] to remove systematic errors in the estimation of the left ventricular volume. Our future research will aim at assessing the impact of our correction technique on the accuracy of ventricular volumes and ejection fraction as estimated by coupled Active Appearance Models in 3D.

Finally, Figure 12.10 shows qualitatively that corrected images are clearly more well-suited for true, three-dimensional exploratory analysis and leads to much improved anatomical understanding.

12.6 Conclusion

A rapid and unsupervised method for retrospectively estimating respiratory-induced slice displacements in four-dimensional cardiac MRI has been described and evaluated. This was presented along with a simple scheme for producing corrected images based on the estimated displacements. Merits and demerits of both have been discussed and potential weaknesses have been pinpointed along with suggestions for future work.

Qualitatively and quantitatively validation carried out on 12 subjects showed that the estimation method was able to deliver a significant improvement over the original images. Joint estimation and correction took less than 10 seconds per subject.

Acknowledgements

M.D. Dorthe Pedersen, M.D., Ph.D. Jens Chr. Nilsson and M.D., Ph.D. Bjørn A. Grønning, Danish Research Centre for Magnetic Resonance, DRCMR, Hvidovre University Hospital are gratefully acknowledged for providing MR images and ground truth two-dimensional myocardial contours for this study.

CHAPTER 13

Bi-temporal 3D Active Appearance Modelling with Applications to Unsupervised Ejection Fraction Estimation from 4D Cardiac MRI

Mikkel B. Stegmann

Abstract

Rapid and unsupervised quantitative analysis is of utmost importance to ensure clinical acceptance of many examinations using cardiac magnetic resonance imaging (MRI). We present a framework that aims at fulfilling these goals for the application of left ventricular ejection fraction estimation in four-dimensional MRI. The theoretical foundation of our work is the generative two-dimensional Active Appearance Models by Cootes et al., here extended to bi-temporal, three-dimensional models. Further issues treated include correction of respiratory induced slice displacements, systole detection, and a texture model pruning strategy. Cross-validation carried out on clinical-quality scans of twelve volunteers indicates that ejection fraction and cardiac blood pool volumes can be estimated automatically and rapidly with an accuracy on par with typical inter-observer variability.

13.1 Introduction

Cardiovascular disease (CVD) continues to be the world's leading cause of death [100]. Consequently, efforts to improve on cardiovascular diagnosis and therapy are not only desirable, but imperative to ensure the public health and a realistic economy within the health sector.

The recently Nobel Prize awarded magnetic resonance imaging (MRI) technique has been firmly established as the gold standard for cardiac imaging. As such, it accurately depicts cardiac structure, function, perfusion and myocardial viability with a capacity unmatched by any other single imaging modality [86]. However, while MRI is a widely accepted method for cardiovascular examinations within research, its use in clinical practice is limited [149]. One major reason for this disparity is the often prohibitive amount of manual interaction required to extract functional indices, see e.g. [99, 145, 161].

This work addresses the crucial problem of estimating global cardiac function in the form of left ventricular (LV) volume, mass and ejection fraction (EF)[1]. A prerequisite for doing so, is an accurate delineation of the myocardium from surrounding tissue, air and blood as shown in a single short-axis cardiac MRI slice in Figure 13.1 (left). Compiling such planar delineations yields the two three-dimensional endocardial (inner) and epicardial (outer) surfaces shown in Figure 13.1 (right) at the end-diastolic phase.

Delineating the myocardium in clinical quality MRI is often not a simple task since many factors during the MR image acquisition deteriorate image quality. These include resolution/time trade-offs, field inhomogeneity, susceptibility arte-

[1]The proportion of blood in the left ventricle that is ejected with each heartbeat.

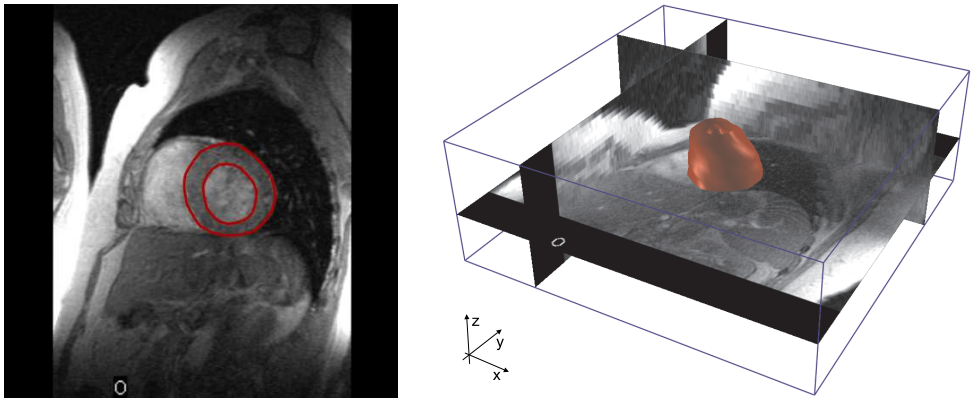


Figure 13.1: Example delineation of the left ventricle in cardiac MRI (left). Four-dimensional cardiac cine image shown at end-diastole. Ground truth endo- and epicardial surfaces shown as shaded meshes (right).

facts, flow artefacts, motion artefacts, partial volume effects, chemical shift artefacts, et cetera. Figure 13.2 shows clearly that models with weak priors – such as local edge detection – may succeed in some, limited parts of the myocardium, but are most likely to utterly fail. Only strong priors can redeem this.

In the following, we will apply global models of shape and image appearance constrained to synthesise plausible instances of the myocardium and the left ventricular blood pool. Solutions will thus be imputed by these models in areas exhibiting weak image evidence, e.g. at parts of the blood/muscle interface. This is analogous to the typical image annotation process where the operator integrates knowledge about the current image slice, not only from the neighbouring slices, but also from previously annotated patients. In fact, these global prior models are built from representative examples containing solutions to the problem at hand; delineation of the myocardium. This powerful concept of learning-based models has recently gained much attention in the image analysis literature due to its virtues of being general computational frameworks that – once trained – remain specific to the domain problem. One widely-used two-dimensional realisation of such a learning-based model is the class of Active Appearance Models (AAMs) [52, 88], which were recently extended to three dimensions [161]. The work presented in the following continues this development with extensions to bi-temporal, three-dimensional Active Appearance Models with applications to estimation of left ventricular volumes and ejection fraction from four-dimensional cardiac cine MRI.

13.2 Data Material

Dynamic and volumetric magnetic resonance images in four dimensions were acquired from twelve healthy, obese (*body mass index* $> 33\ \mathrm{kg/m^2}$) subjects using a whole-body MR unit (Siemens Impact) operating at 1.0 Tesla. The pulse sequence was an ECG-triggered, breath-hold, fast low angle shot (FLASH) cinematographic pulse sequence. Matrix=256×256, field of view=263×350 mm, slice thickness=6 mm, inter slice gap=0 mm, phase time=55 ms. The spatio-temporal volumes had an average size of $(x, y, z, phases)$ = 256×256×23×12 voxels, each stored in 16 bit.

The endocardial (endo) and epicardial (epi) contours of the left ventricle were annotated at the end-diastole (ED) and the end-systole (ES) by manually placing landmarks along the myocardial border, see Figure 13.1 (left). The overall image quality in this data set is illustrated in Figure 13.2.

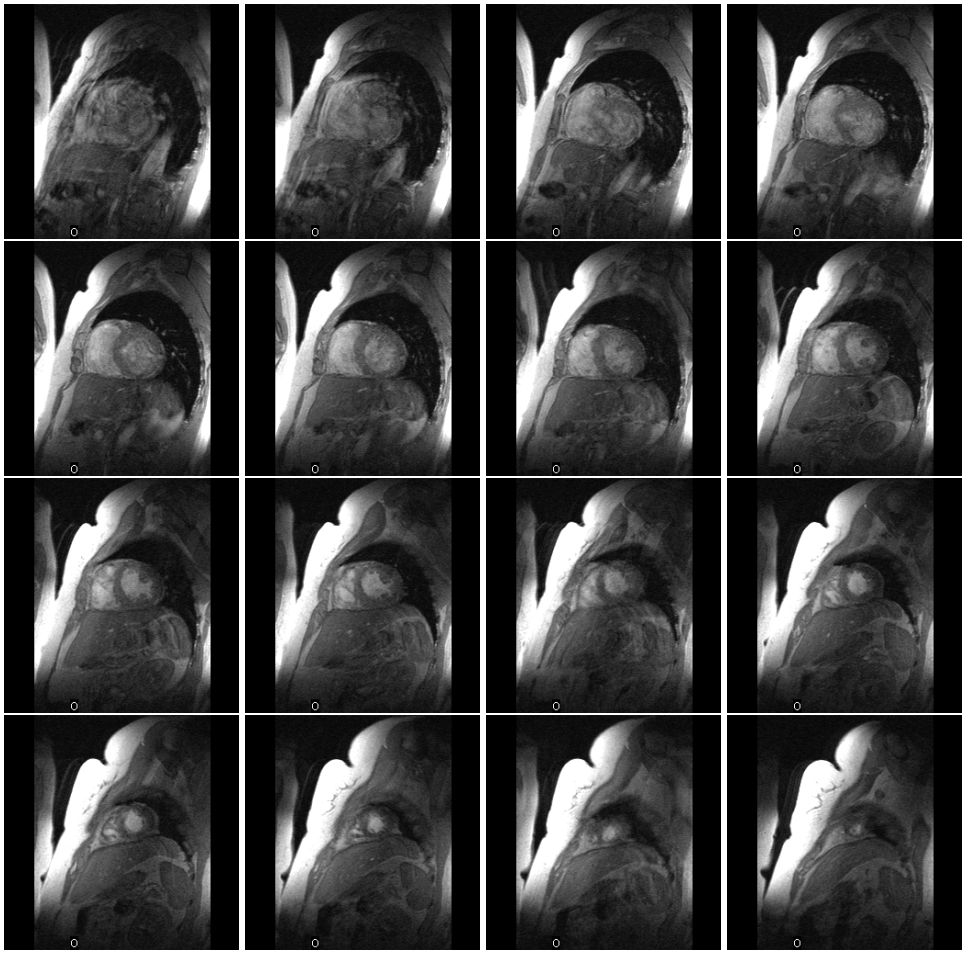


Figure 13.2: Typical image quality; 16 of 21 end-diastolic slice images shown in rows, from base to apex for patient 1. The dark area predominately in the first image quadrant is the lung.

13.3 Methods

The methodological overview presented in this section is given in two parts. First, the Active Appearance Models (AAMs) are presented using broad terms, without touching the question of dimensionality or the domain problem in itself. Then specific details regarding modelling of bi-temporal cardiac data are given; with emphasis on issues pertinent to the extension of AAMs to higher dimensions than two.

13.3.1 Active Appearance Models

Interpretation by synthesis has been shown to be a very powerful approach to image analysis. On a coarse level, this approach consists of two parts; i) a mathematical model, which is able to mimic the image formation process, and ii) a search regime, which is able to match this model to an image by letting the image synthesised by the model – in some sense – be "similar" to the unseen image.

For such a strategy to be useful, certain requirements must be met for both parts. The model must either encode specific features of interests and/or contain parameters that reflect interesting latent variables of the objects in question. For example this could be point features locating interesting structures, and a variable related to the age of the object. Additionally, this model should only be capable of synthesising valid image instances of the object class in question. Matching such a model to an unknown image would thus draw inference about these properties. The matching procedure should provide a meaningful model-to-image match in a reasonable amount of time, dependent on the application, or alternatively reject the presence of the sought-after object.

Active Appearance Models (AAMs) [52, 88] represent one method that has gained considerable attention in the literature, seeking to meet the above requirements. The encoded properties of the model are correlated movement of landmark points and latent texture variables controlling image appearance normalised for shape changes given by landmark points. In essence a fruitful marriage between the ideas of Eigenface Models [237], Point Distribution Models [57] and Active Shape Models [67] with some ingenious refinements. All these techniques originated in the early nineties. Since their introduction in 1998, AAMs have been refined and adapted in many aspects, see e.g. [54, 65]. A recent summary of AAM refinements and applications within medical imaging can be found in [213].

Formally, AAMs establish a compact parameterisation of object variability, as learned from a representative training set. Objects are defined by marking up each example with points of correspondence (i.e. landmarks) over the training set either by hand, or by semi- to completely automated methods. Using a learning-based optimisation strategy, AAMs can be rapidly fitted to unseen images.

Variability is modelled by means of a Principal Component Analysis (PCA), i.e. an eigenanalysis of the dispersions of shape and texture. Let there be given Q training

examples for an object class, and let each example be represented by a set of N landmark points and M texture samples. Shape examples are aligned to a normalised common mean using a Generalised Procrustes Analysis [108]. Texture examples are warped into correspondence by forcing landmarks to match, intensity normalised, and subsequently sampled from this *shape-free* reference. Typically, this geometrical reference frame is the Procrustes mean shape. Let $\mathbf{s}$ and $\mathbf{t}$ denote a synthesised shape and texture and let $\bar{\mathbf{s}}$ and $\bar{\mathbf{t}}$ denote the corresponding sample means. New instances are now generated by adjusting the principal component scores, $\mathbf{b}_s$ and $\mathbf{b}_t$ in

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{\Phi}_s \mathbf{b}_s \tag{13.1}$$

and

$$\mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_t \mathbf{b}_t \tag{13.2}$$

where $\mathbf{\Phi}_s$ and $\mathbf{\Phi}_t$ are eigenvectors of the shape and texture dispersions estimated from the training set. To obtain a combined shape and texture parameterisation, $\mathbf{c}$, the values of $\mathbf{b}_s$ and $\mathbf{b}_t$ over the training set are combined into

$$\mathbf{b} = \begin{bmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_t \end{bmatrix} = \begin{bmatrix} \mathbf{W}_s \mathbf{\Phi}_s^{\mathrm{T}} (\mathbf{s} - \bar{\mathbf{s}}) \\ \mathbf{\Phi}_t^{\mathrm{T}} (\mathbf{t} - \bar{\mathbf{t}}) \end{bmatrix}. \tag{13.3}$$

A suitable weighting between pixel distances and pixel intensities is carried out through the diagonal matrix $\mathbf{W}_s$. To recover any correlation between shape and texture the two eigenspaces are usually coupled through a third PC transform,

$$\mathbf{b} = \mathbf{\Phi}_c \mathbf{c} = \begin{bmatrix} \mathbf{\Phi}_{c,s} \\ \mathbf{\Phi}_{c,t} \end{bmatrix} \mathbf{c}, \tag{13.4}$$

obtaining the combined appearance model parameters, $\mathbf{c}$, that generate new object instances by

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{\Phi}_s \mathbf{W}_s^{-1} \mathbf{\Phi}_{c,s} \mathbf{c} \tag{13.5}$$

and

$$\mathbf{t} = \bar{\mathbf{t}} + \mathbf{\Phi}_t \mathbf{\Phi}_{c,t} \mathbf{c}. \tag{13.6}$$

The object instance, $\{\mathbf{s}, \mathbf{t}\}$, is synthesised into an image by warping the pixel intensities of $\mathbf{t}$ into the geometry of the shape $\mathbf{s}$ and applying the current pose parameters; translation, rotation, and shape size denoted by $\mathbf{p}$.

An AAM is matched to an unseen image using a least-squares criterion by an iterative updating scheme, which is based on a fixed Jacobian estimate [54], or originally, a principal component regression [52]. For this work we have used the former approach, treated in further detail elsewhere [213]. In short, this allows model parameters, $\{\mathbf{c}, \mathbf{p}\}$, to be updated based on the difference between the model texture, $\mathbf{t}$, and image texture, $\mathbf{t}_{image}$, the latter being the image texture covered by the model. Let $\mathbf{\Psi}$ denote a fixed update matrix pre-calculated from the training set. Model parameter updates are then estimated by

$$[\, \delta\mathbf{c}^{\mathsf{T}} \; \delta\mathbf{p}^{\mathsf{T}} \,]^{\mathsf{T}} = \mathbf{\Psi}(\mathbf{t}_{image} - \mathbf{t}). \tag{13.7}$$

For further details on AAMs refer to [52, 54, 65] (and possibly [203] as a supplement).

13.3.2 Bi-temporal Cardiac Modelling in Three Dimensions

Originally, AAMs operated in the two-dimensional domain only. However, a study employing a truly three-dimensional AAM has recently been presented [144, 161, 162] ([161] being the most comprehensive reference). Semi three-dimensional applications have been described in [19, 20, 35].

While bearing resemblance to the work on 3D cardiac AAMs presented by Mitchell et al. [161], this study differs in several aspects. Partly, due to the coupled nature of the application and partly due to differences in design and usage of auxiliary methods. We have sought to make this explicit in following with some mathematical details placed in appendices for completeness. Reproduction of the presented method should be possible from this reference and a detailed description of the 2D AAM as referenced in Section 13.3.1.

Correction of Respiratory Induced Slice Displacements

Prior to landmarking and subsequent analysis, 4D cardiac cine MRI should optimally be corrected for motion induced by respiration during the acquisition process. This motion is due to the 4D image being constructed from a set of 2D slice image time-series, each having a potentially different inspiration depth, which causes a displacement of the upper abdomen and the heart.

This correction can be carried out by unsupervised image analysis. Statistical models of shape and appearance in 2D are employed to localise the left and right ventricle in a central slice between the apex and base. The segmented region is subsequently propagated through all slices in the apex-basal direction using a non-rigid registration process, thus providing an estimate of the inter-slice translation due to variation in inspiration depth for each breath-hold. The details of this method alongside a quantitative and qualitative validation can be found in [225]. In summary, we found that it was possible to markedly suppress respiratory motion artefacts without manual interaction using less than ten seconds of CPU time for each 4D image. An example result is shown in Figure 13.3 for patient 4. Notice the obvious reconstruction of cardiac and upper abdominal structures in the left half of Figure 13.3 (right), while the respiratory-displacements have been ”transferred” to the structures not affected by respiration, i.e. right half of Figure 13.3 (right).

Point Correspondences on the Myocardium

At the heart of shape modelling lies the requirement of *correspondence*. Which, in the present case, translates to a demand of tissue correspondences on the myocardium;

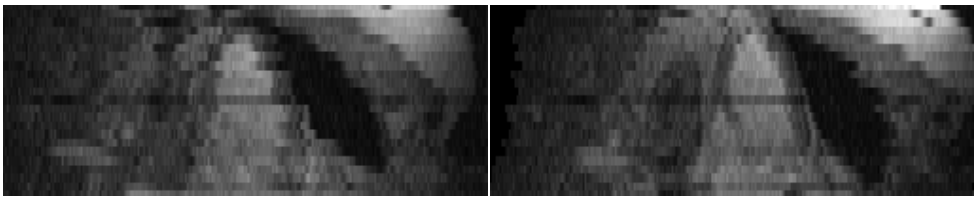


Figure 13.3: Unsupervised correction of respiratory induced slice displacements. Original yz-plane of patient 4 (left). Corrected yz-plane of patient 4 (right).

a near landmark-less structure. Due to the simple geometry of the endocardial and epicardial contours, inter-slice correspondences were established by arc-length resampling. The starting point was placed at the anterior junction between the left and right ventricle. Sampling density was chosen to let the endocardial and epicardial surfaces be approximately equally dense on average. The number of sampling points was held constant for all slices. Subsequently, z-axis correspondence was established by linear interpolation of corresponding points between slices. All volumes were resampled to have the average number of slice contours. This inter-slice correspondence enabled a simple conversion from the set of contours to endocardial and epicardial surfaces in a meshing process connecting corresponding points to form strips of triangles, for which surface normals could be estimated.

This crude approach to landmarking, is justified by the combination of scan type (short-axis), the simple geometry of the myocardium, and finally by the low-frequency contours obtained using the common clinical practice (see e.g. [161, 177]) of excluding papillary muscles and ventricular trabeculae from the endocardial contour. More sophisticated approaches to cardiac landmarking were presented in [98, 101] and [153]. Another recent landmarking method based on minimisation of description length [75, 77] is also of interest to this problem, although it currently only handles three-dimensional objects with topology equivalent to a sphere.

Coupling End-diastole and End-systole

To model coherence between the two cardiac phases of interest; the end-diastole and end-systole, a coupled AAM representation is chosen. This was first explored in [71] for interpretation of multi-view face images. Later [159] coupled cardiac time-sequences, an approach which was also pursued in [33, 36, 37, 146, 198] and for long and short axis MRI in [147]. However, most similar to our approach is the recent AAM work presented in [174, 175], where two-dimensional, end-diastolic and end-systolic angiograms were coupled in shape and appearance, but not in pose due to the nature of the data material. However, in our case, the main objective is to constrain the differences in pose by encoding these into the shape model. This is accomplished by a simple concatenation shown below.

Let a shape consisting of N points in three dimensions be defined by a $3N$ vector,

$$\mathbf{s} = [\ x_1\ y_1\ z_1\ x_2\ y_2\ z_2\ \ldots\ x_N\ y_N\ z_N\ ]^{\mathsf{T}}. \tag{13.8}$$

One combined patient ED-ES observation is then composed of three such shapes; $\mathbf{s}_{endo_{ED}}$, $\mathbf{s}_{epi_{ED}}$ and $\mathbf{s}_{endo_{ES}}$ by concatenation prior to Procrustes alignment and tangent space projection,

$$\mathbf{s}_{subject} = \begin{bmatrix} \mathbf{s}_{endo_{ED}} \\ \mathbf{s}_{epi_{ED}} \\ \mathbf{s}_{endo_{ES}} \end{bmatrix}. \tag{13.9}$$

Likewise, the associated image texture is composed from samplings of two 3D images; the volume spanned by the blood pool and the myocardium from the end-diastolic phase, $\mathbf{t}_{ED}$, and the blood pool of the end-systolic phase, $\mathbf{t}_{ES}$. The combined texture vector then becomes

$$\mathbf{t} = [\ \mathbf{t}_{ED}^{\mathsf{T}}\ \mathbf{t}_{ES}^{\mathsf{T}}\ ]^{\mathsf{T}}. \tag{13.10}$$

Procrustes Alignment

Building a model of shape variability as in AAMs, requires per se elimination of effects stemming from translation, rotation and isotropic scaling. This is accomplished by an Ordinary Procrustes Analysis (OPA) bringing two shapes into alignment, or by a Generalised Procrustes Analysis (GPA) aligning a set of shapes to their common mean. Unlike the impression given in the earlier 3D and semi-3D AAM work [35, 161, 162], well-behaved solutions to both problems do exist in two and three dimensions formulated in terms of orthonormal matrices, see [7, 80, 120, 231].

In summary, closed-form OPA solutions exist in 2D and 3D, while a closed-form GPA solution only exists in 2D. In the current work, we employ OPA in 3D based on a singular value decomposition (SVD) of the inter-shape correlation matrix, avoiding the quaternion formulation used in [35, 161, 162] altogether. We obtain the GPA solution by a simple iterative scheme, see e.g. [31, 65]. The mathematical details of OPA in k dimensions are given in Appendix 13.A.

Texture Correspondences

Two-dimensional image warping using a triangular mesh is a fast – and often sufficient – solution to the ill-conditioned problem of establishing a dense planar correspondence from two sparse sets of corresponding points. Typically, a Delaunay triangulation of one of the point sets is used to form this mesh of two-dimensional simplexes. See e.g. [65, 203] for the details of this warping process that also can be accelerated using contemporary graphics hardware [213].

In three dimensions, the Delaunay decomposition produces three-dimensional simplexes, called tetrahedra. Alternatively, these tetrahedra can be produced by a dedicated hand-tailored method as in done in [161]. To avoid this tedious process, we employ the freely available 3D Delaunay implementation TetGen [195] to produce a tetrahedral decomposition of the mean shape. This ensures a rapid and unsupervised model building process, which is adaptive to the mean geometry, $\bar{\mathbf{s}}$, encountered in a given domain problem.

Let $\mathbf{b}_i = [\, \alpha \, \beta \, \gamma \, \delta \,]^\mathsf{T}$ denote the barycentric coordinates[2] of a tetrahedron given by $\mathbf{T}_i$. Then, let $\mathcal{T} = \{\mathbf{T}_i\}_1^R$ denote the set of R tetrahedra in the mean configuration (vertices incident to the mean shape) and let $\mathcal{T}' = \{\mathbf{T}'_i\}_1^R$ denote a corresponding set of deformed tetrahedra. Using the notation of Appendix 13.B, warping a point from $\mathcal{T}$ to $\mathcal{T}'$ can then be given by

$$\mathbf{p}'_i = \mathcal{W}(\mathbf{p}_i) = \mathcal{W}(\mathbf{T}_i \mathbf{b}_i) = \mathbf{T}'_i \mathbf{b}_i. \tag{13.11}$$

Consequently, during model generation, a set of texture sample positions of the mean tetrahedral configuration are chosen, $\{\mathbf{p}_i\}_1^M$, and their corresponding barycentric coordinates, $\{\mathbf{b}_i\}_1^M$, are determined (using Appendix 13.B) and stored. Texture sample positions for arbitrary tetrahedral configurations during AAM search are now obtained by simple matrix-vector multiplications.

In the current work, texture samples are distributed equidistantly with a spacing of one voxel within the axis-aligned bounding box containing the mean shape. This involves solving the point-in-tetrahedra problem; a task that can be substantially sped up by a spatial search tree of the tetrahedra. However, being an off-line process, a (slow) linear search using an axis-aligned bounding box test for each tetrahedron, followed by the exact test in Appendix 13.B, was sufficient in our case.

A computationally more demanding alternative to the above could be to choose the three-dimensional warp field that minimises the bending energy (the integral over $\mathbb{R}^3$ of the squares of the second derivatives). This solution is often referred to as thin-plate splines, see e.g. [30].

Lastly, notice that in the bi-temporal case, two separate warping functions must be established from two tetrahedral decompositions; one for end-diastole and one for end-systole.

Texture Sampling

As stated above, all elements in $\{\mathbf{p}_i\}_1^M$ remain fixed and – to ease synthesis of model textures – the values of these are chosen to take on integer values forming a regular lattice. Elements of $\{\mathbf{p}'_i\}_1^M$ will thus typically be irrational numbers for which no image data exist. Consequently, an interpolation method is required. For computational efficiency, this work uses the commonly used tri-linear interpolation (the volumetric equivalent of bi-linear interpolation in 2D).

[2] See Appendix 13.B.

Alternatively, a more optimal reconstruction could be obtained by sinc interpolation, which unfortunately exhibits excessive computational demands compared to tri-linear interpolation. However, it was recently shown that the major error component from premature truncation of the sinc kernel can be compensated for by using a renormalisation strategy [232]. Alternatively, sophisticated adaptive filtering strategies can be taken. See e.g. the impressive results on interpolation of anisotropic voxel data from computed tomography in [248].

Proximity Modelling

In medical images, organs and structures typically have a very homogenous appearance. Since AAMs only model the appearance spanned by the triangular (2D) or tetrahedral (3D) mesh, no information regarding the surroundings is modelled if the outer landmarks are placed exactly on the object boundary. This is naturally typically the case, as this is the desired structure the model should infer knowledge about.

The two above issues lead to models, which can show minima in the model-to-image cost function when the outer surface (or contour in 2D) is erroneously inside the object of interest. In the present case, the epicardial surface could be between the blood pool and the lung. Therefore, we apply a simple scheme for modelling the texture variation of the object proximity. This is accomplished by adding surface normals outwards from the shape. These normals are denoted *whiskers* and added implicitly during texture sampling with a scale relative to the current shape size. Texture samples obtained by sampling along whiskers are now concatenated to the texture vector of the conventional AAM texture samples.

Texture Model Pruning

AAMs build on assumptions of normally distributed data. Consequently, the seemingly absence and presence of papillary muscles are per se not well modelled in the texture model. Therefore we propose a pruning strategy, which excludes these problematic texture sample positions. In lieu of tedious manual pruning; a "suitability" measure, $\mathcal{S}(t_i)$, is used to obtain such pruning.

Optimally, $\mathcal{S}$ should measure the non-Gaussianity; a striking resemblance to a problem occurring in the estimation of independent components in ICA (Independent Component Analysis). In [126] this is solved by exploiting a central result from information theory, namely that a Gaussian distributed random variable exhibits maximal entropy among all random variables of equal variance. This leads to the concept of *negentropy*, that [126] presents computationally feasible estimators for, which are far more robust than the traditional kurtosis measure.

However, due to the small sample size of our data (twelve samples per element in $\mathbf{t}$), texture models are pruned on the basis of variance; $\mathcal{S}(t_i) = \sum_{j=1}^{Q}(t_{i_j} - \bar{t}_{i_j})^2$.

All texture sample positions above a user-selected percentile of $\mathcal{S}$ are now excluded from the training set texture samples and omitted in future samplings during image search.

Pose Representation

Pose in three dimensions can be parameterised by seven components,

$$\mathbf{p} = [\, t_x \; t_y \; t_z \; \theta_x \; \theta_y \; \theta_z \; 1 - s \,], \tag{13.12}$$

denoting spatial translation, rotation and one minus the isotropic scaling, respectively. Having the identity transformation represented by the null vector, this representation is suitable for representing pose displacements in the aligned shape domain employed by AAMs. Let us denote this domain $\mathcal{A}$ and let the image shape domain – where landmarks are given in terms of voxel positions – be denoted by $\mathcal{I}$. A convienent representation of the mapping $\mathcal{I} \mapsto \mathcal{A}$ is

$$\mathcal{Q} = \{\, t_x,\; t_y,\; t_z,\; s,\; \mathbf{R} \,\} \tag{13.13}$$

where $\mathbf{R} \in \mathbb{R}^{3\times3}$ is an orthogonal matrix rotating a shape in $\mathcal{I}$ to the reference orientation in $\mathcal{A}$ as obtained by OPA. Shape displacement experiments required during the estimation of $\mathbf{\Psi}$ in the AAM building process can now be composed by $\mathbf{p} * \mathcal{Q}$, i.e. a shape displacement in $\mathcal{A}$ and a mapping (fixed for each training shape) to $\mathcal{I}$.

Systole Detection

To render ejection fraction estimation using a bi-temporal 3D model completely unsupervised, the systolic phase needs to be estimated. Due to the flexibility of the model, this can be straightforward estimated by running several searches on different time phases near the central part of the time-series and choose the search result maximising the difference between the endocardial volumes (i.e. the stroke volume);

$$\hat{t}_{systole} = \arg\max_t \left(\mathrm{Volume}\{\mathbf{s}_{endo_0}\} - \mathrm{Volume}\{\mathbf{s}_{endo_t}\} \right). \tag{13.14}$$

13.4 Implementation

The presented bi-temporal 3D AAM framework was implemented in standard C++ using three major third-party libraries. Singular value decomposition and eigenvalue solver were provided by LAPACK [4]. A fast, architecture-specific BLAS implementation for matrix-vector multiplication was provided by Intel Math Kernel Library [127]. Three-dimensional Delaunay decomposition was provided by TetGen [195]. In addition to the AAM implementation, a framework for data visualisation in four dimensions were written in C++ using OpenGL [93]. This produced all figures in this article showing MRI data or geometry.

To obtain a reasonable level of performance, aggressive C++ *inlining* has been used; in particular for the warping code. Computational complexity and cache consistency were considered during design and implementation, but neither code nor memory optimisation has been applied in the current implementation.

13.5 Experimental Results

To assess the ability of the described AAM framework to locate the left ventricle in 4D cine MRI, a series of cross-validation experiments were carried out on a 2.4 GHz Pentium PC equipped with 1 GB RAM. Due to the small sample size (twelve patients), a leave-one-out validation approach was chosen. Hence, twelve models were built on eleven training examples and the remaining example was used for testing. For each experiment, the bi-temporal AAM searched the end-diastolic phase and five central frames to locate the end-systole. In each of these five searches, the AAM was initialised in a 3×3 grid in the xy slice plane around the mean pose configuration as determined from the (independent) training set. The grid spacing used was 33% of the mean shape width in x and y, respectively. This grid size showed sufficient coverage of the pose variation in the current training set, since at least one of the nine searches converged onto the true location of the heart in all patients. For other data sets, a different grid size may be needed. In total, $12 \times 5 \times 9 = 540$ AAM searches were carried out.

All shape models contained 1560 points, i.e. 4680 variables. To reduce memory requirements and speed up computations, all input images were subsampled in x and y to 128×128, z remained unchanged. Texture models contained approximately 22000 texture samples. The proximity of the left ventricle was modelled by adding whiskers at each landmark with a length of five voxels (measured on the mean shape sized to mean size). Five samples for each whisker were acquired during texture sampling. The texture model pruning retained the 70% samples showing the least variance over the training set. Subsequently, shape and texture PCA models were truncated to explain 98% variation in the training set (see e.g. [213]). To gain increased flexibility from the limited training set, shape and texture eigenspaces were not coupled, i.e. $\mathbf{\Phi}_c = \mathbf{I}$.

Table 13.1 shows the signed and unsigned mean errors, between AAM and the human observer, of the end-diastolic volume (EDV), end-systolic volume (ESV) and ejection fraction (EF). A more detailed impression of agreement between these two methods for estimation of global functional cardiac indices is given by the Bland-Altman plots [26] in Figure 13.4. The accuracy obtained in each patient is illustrated in Figure 13.5.

Qualitative results are given for patient 4 in Figures 13.6–13.9, showing the initial and converged model. This verbosity, in terms of images, is due to the discrepancy in dimensionality between paper and the MRI data. Movie sequences illustrating

Table 13.1: Unsigned and signed errors for the estimated cardiac volumes and ejection fraction using automatic systole detection.

	Unsigned mean errror	Signed mean error
Endo EDV	6.8±3.7 ml	-0.7±7.9 ml
Endo ESV	3.9±2.5 ml	1.3±4.6 ml
Epi EDV	23.8±17.2 ml	5.6±29.7 ml
EF	2.8±1.9 percent	-1.2±3.3 percent

Table 13.2: Unsigned and signed errors for the estimated cardiac volumes and ejection fraction using ground truth systole.

	Unsigned mean errror	Signed mean error
Endo EDV	6.5±5.2 ml	2.5±8.1 ml
Endo ESV	3.8±2.3 ml	0.8±4.5 ml
Epi EDV	23.3±24.8 ml	13.1±32.0 ml
EF	2.1±2.0 percent	0.0±3.0 percent

the model convergence can be obtained electronically by request to the author.[3]

The systole detection determined the time phase given by the human observer in five cases (42%). In six cases (50%) the neighbouring phase was deemed the systole. In one case (8%) the detected systole and the observer-given differed by two phases. All seven systolic phases that did not match the observer-given phase were detected in a later phase. To assess the potential confounding between systole detection and ventricular volume estimation, a set leave-one-out experiments using the ground truth systole were carried out. A summary of these results is given in Table 13.2.

A single leave-one-out experiment peaked with respect to memory usage at 153 MB in this unoptimised implementation. Model building and model search timings are shown in Table 13.3. The figures in the third and fourth row are obtained by dividing the second row with the number of phase searches (five) and initialisation grid sites (nine). One image texture sampling took 15 ms on average.

[3]Or, by visiting http://www.imm.dtu.dk/∼aam/.

Table 13.3: Model building and model search timings using automatic systole detection.

	All experiments [seconds]	Per experiment [seconds]
Model building	571.0	47.6±2.4
Model search	217.3	18.1±3.5
Average single phase search	43.5	3.6
Average single search	4.8	0.4

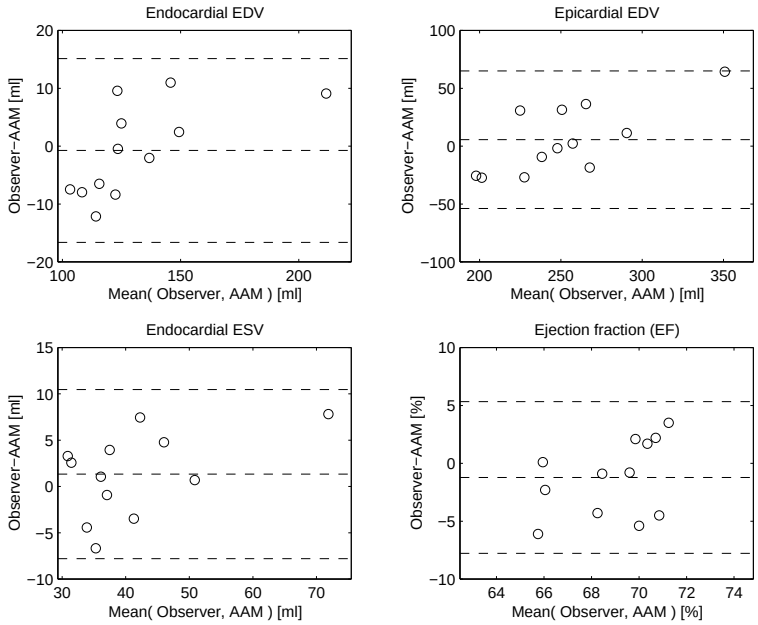


Figure 13.4: Bland-Altman plots for end-diastolic endocardial volume, end-diastolic epicardial volume, end-systolic endocardial volume, and left ventricular ejection fraction. Limits of agreement are plotted as dashed lines at ± 2 standard deviations.

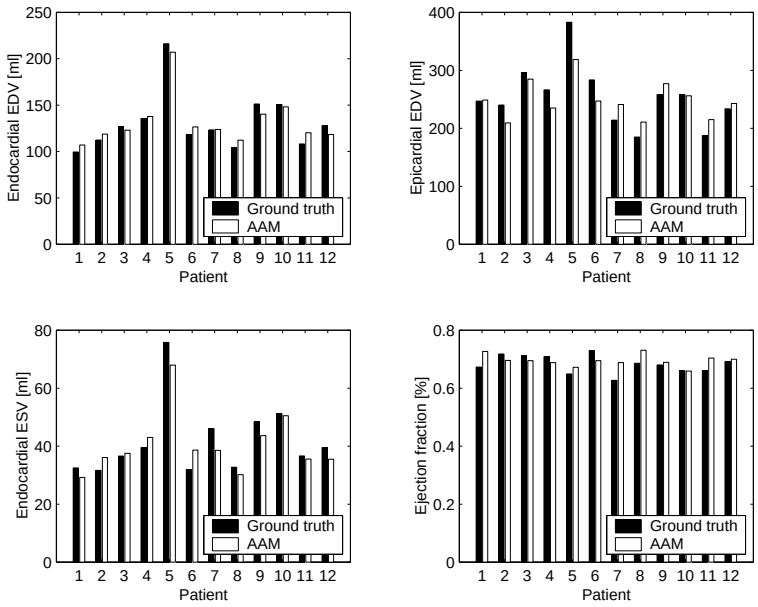


Figure 13.5: Bar plots for end-diastolic endocardial volume, end-diastolic epicardial volume, end-systolic endocardial volume, and left ventricular ejection fraction.

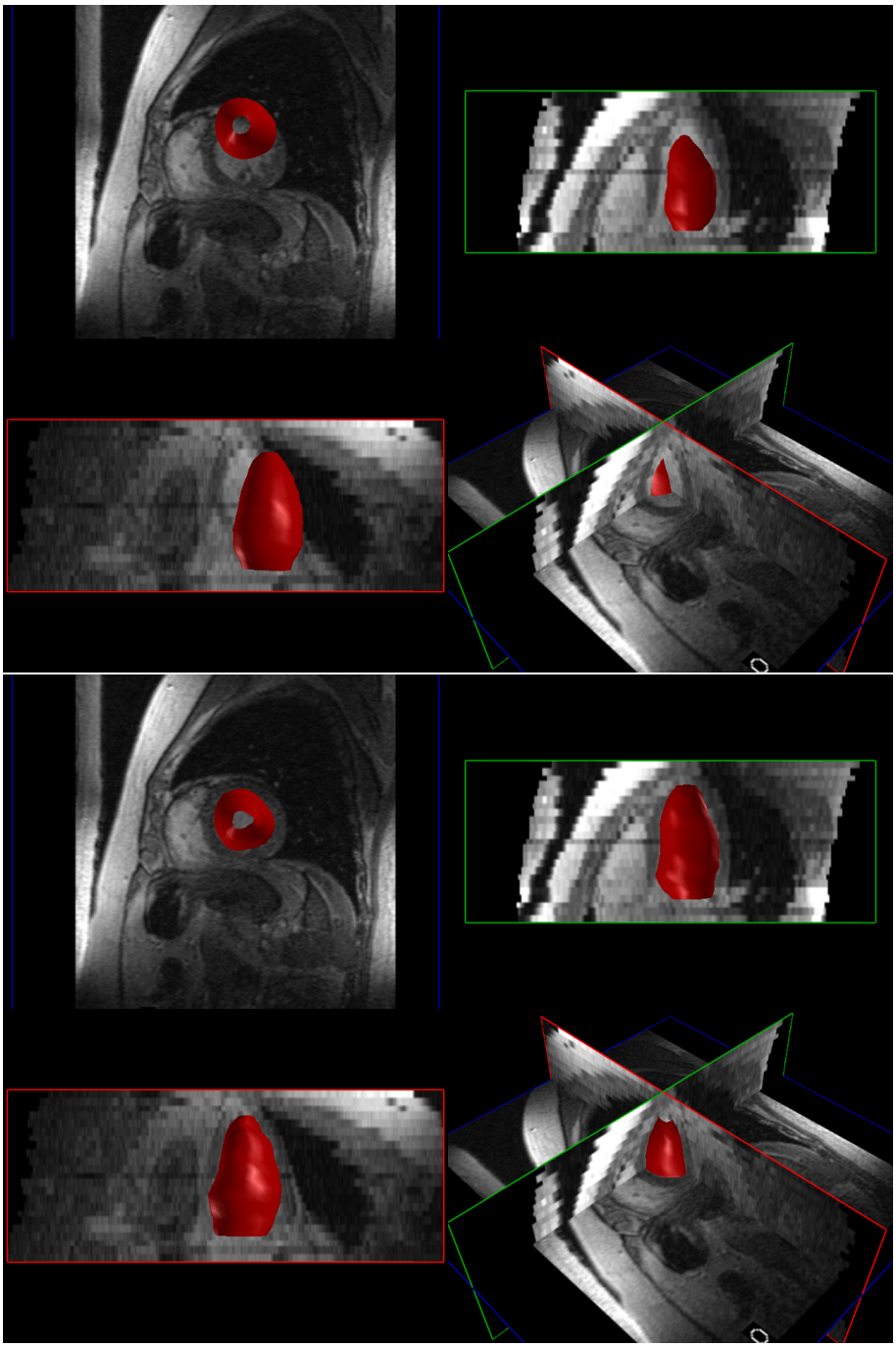


Figure 13.6: End-diastolic endocardial surface shown in the initial and optimal model configuration for patient 4 after 17 iterations.

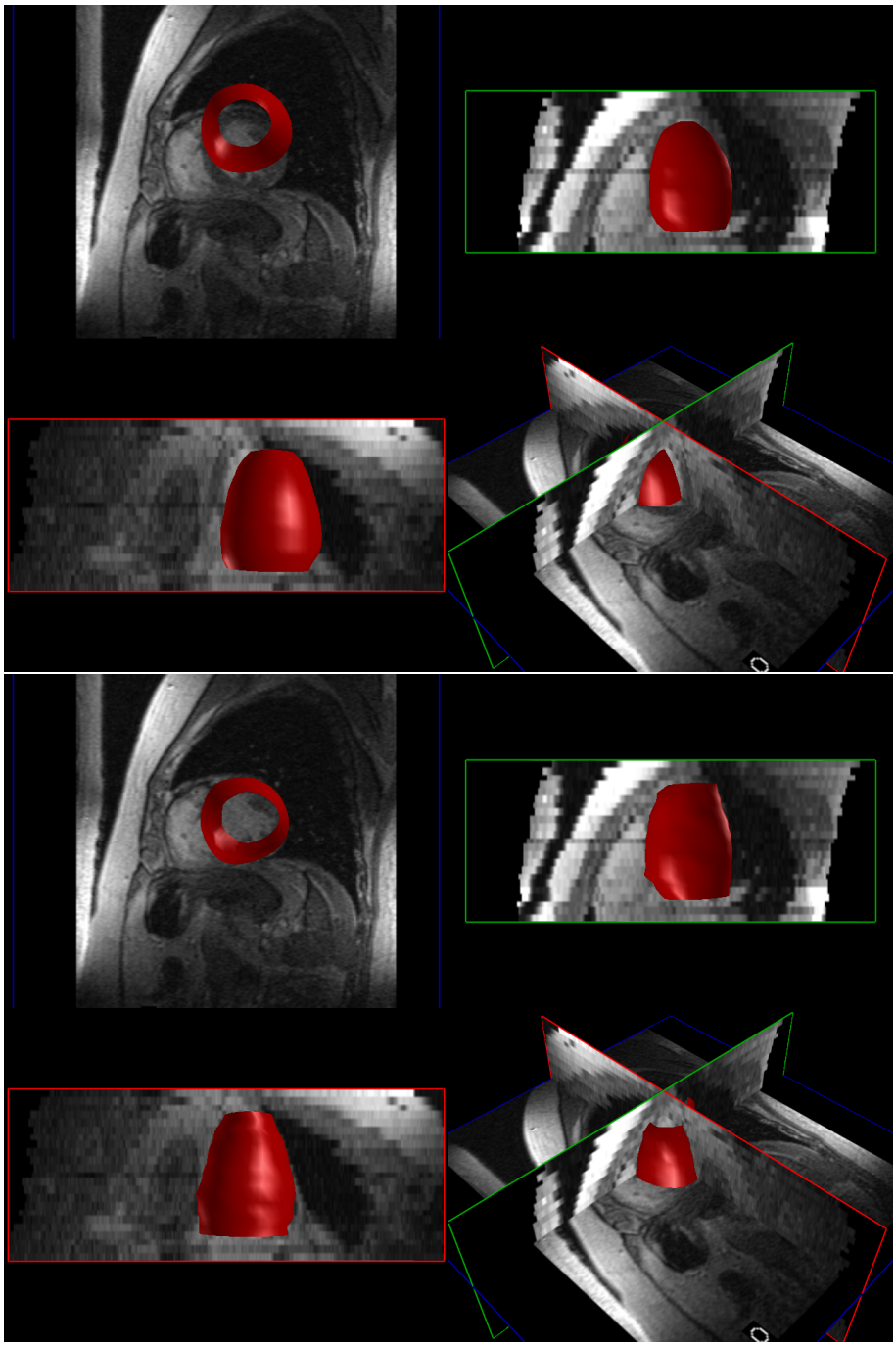


Figure 13.7: End-diastolic epicardial surface shown in the initial and optimal model configuration for patient 4 after 17 iterations.

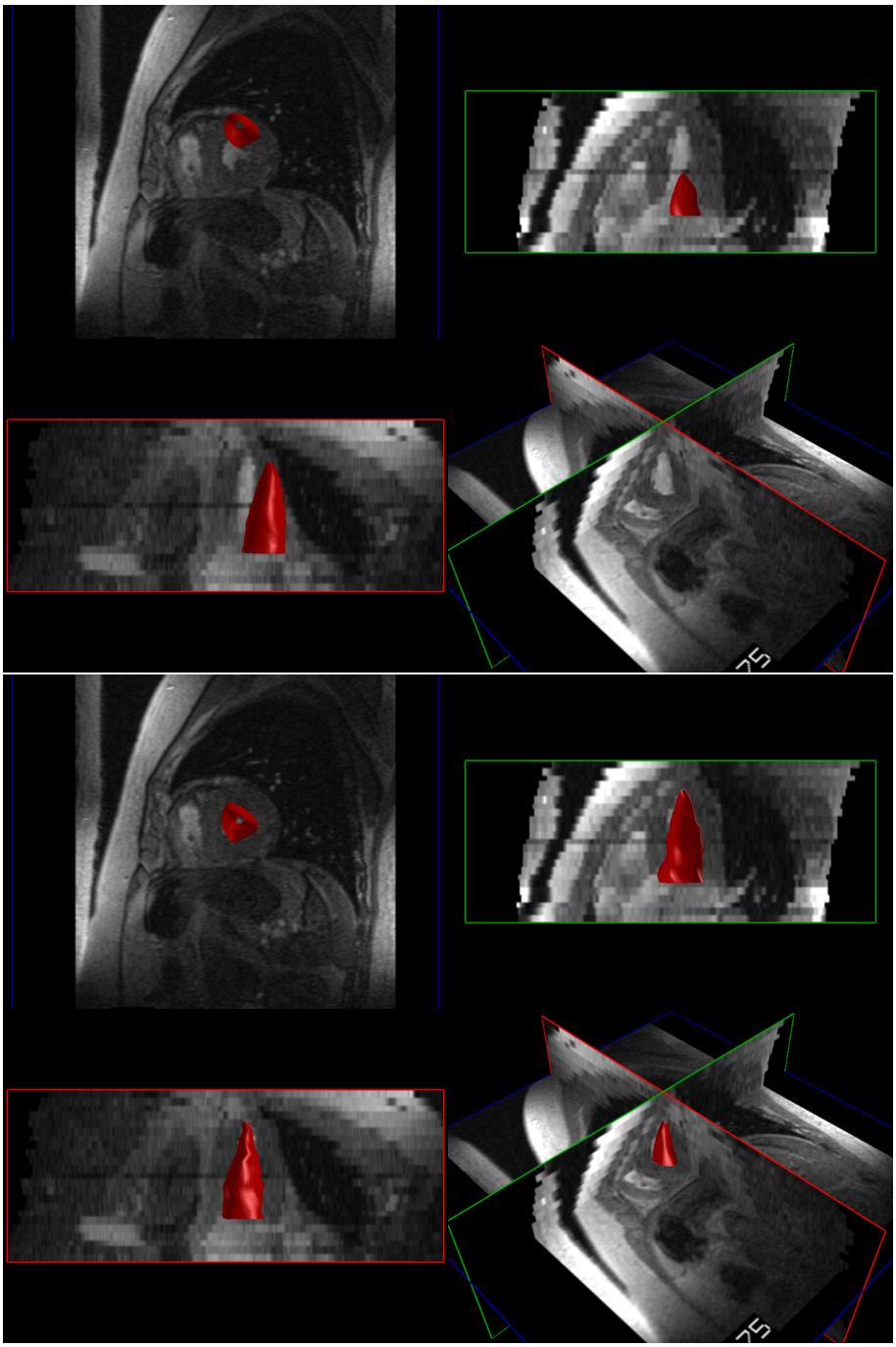


Figure 13.8: End-systolic endocardial surface shown in the initial and optimal model configuration for patient 4 after 17 iterations.

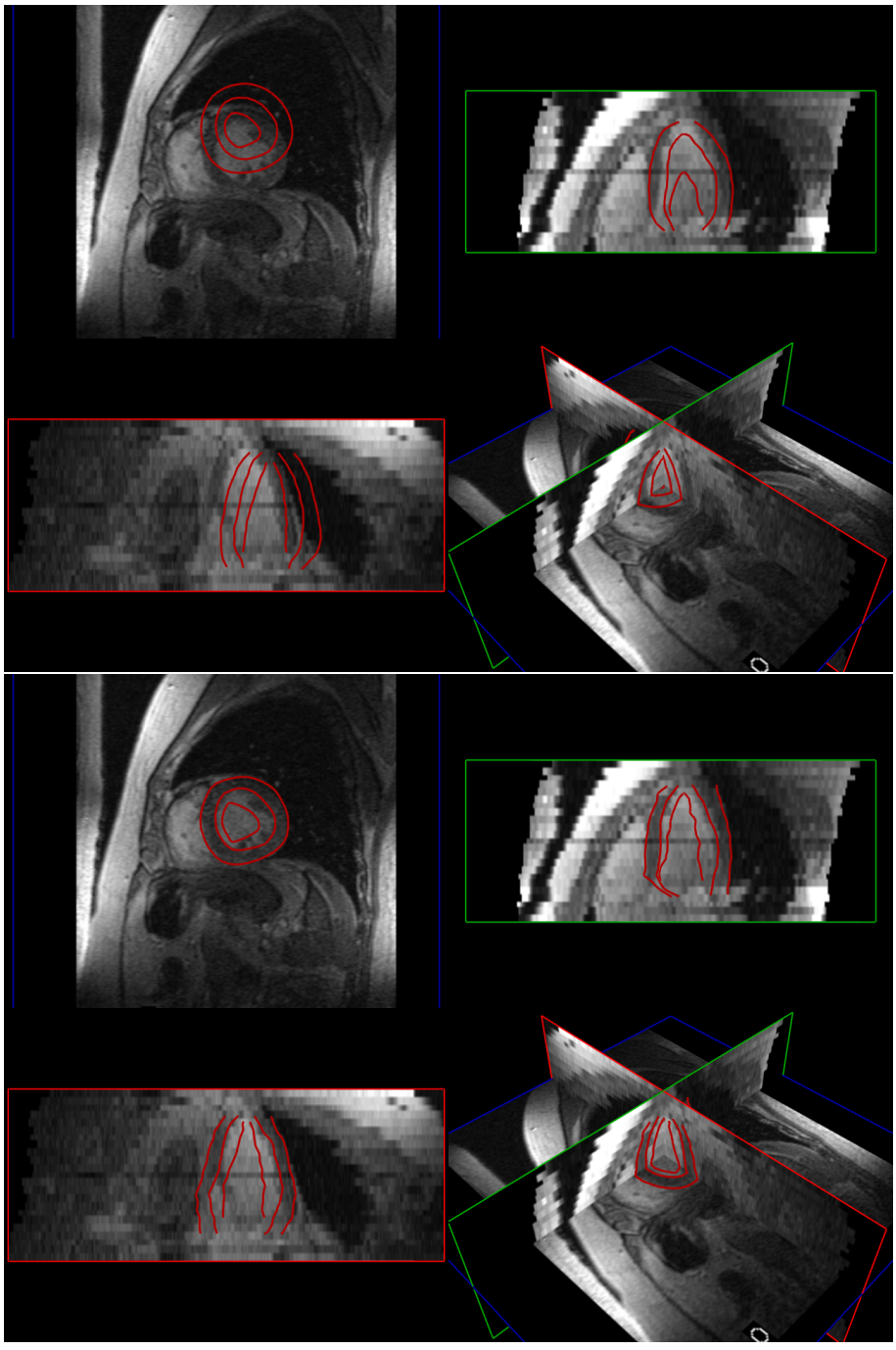


Figure 13.9: End-diastolic end-systolic endocardial and end-diastolic epicardial contours shown in the initial and optimal model configuration for patient 4 after 17 iterations.

In comparison to Table 13.2, a similar set of experiments was carried out, but no pruning of the texture model, no whiskers and using the traditional coupled shape and texture eigenspaces. The unsigned mean errors were: Endo EDV=10.0$\pm$9.1 ml; Endo ESV=5.5$\pm$4.0 ml; Epi EDV=28.6$\pm$23.3 ml; EF 3.2$\pm$1.7 percent.

Repeating the experiments in Table 13.2, but without the described slice correction procedure showed a markedly decrease in accuracy for all figures but the Epi EDV. The unsigned mean errors were: Endo EDV=11.7$\pm$9.3 ml; Endo ESV=6.7$\pm$5.0 ml; Epi EDV=24.6$\pm$17.6 ml; EF 3.1$\pm$2.5 percent.

Finally, to prove that full-resolution modelling of the input data was possible, an additional set of experiments was carried out. These were similar to the ones in Table 13.2, but with no x and y subsampling. Results were comparable, but errors were – although modestly – consistently higher compared to the equivalent subsampled results.

13.6 Discussion

The simple grid-based initialisation strategy succeeded in localising the heart in all twelve patients. This success may in part be attributed to the relatively small training set showing a moderate variation in pose. For larger training sets, it may be needed to perform leave-one-out analyses to determine the limits of convergence for each pose and model parameter. These, in combination with the observed distribution of pose and model parameters, will then determine the required minimal grid size and spacing to guarantee successful generic initialisation. Additionally, a multi-resolution implementation can be used to speed up this process.

Our experiments showed that the end-systole localisation obtained is acceptable. The results in Table 13.1 are – by design – biased towards larger end-diastolic volumes and smaller end-systolic volumes. This, and the fact that only five out of twelve end-systolic phases were correctly classified, are still considered acceptable. The end-systolic phases (and thereby the associated volume) were manually determined from a single slice image only. On the contrary, the model was able to search for a global volumetric minimum. Further, considering the coarse temporal sampling and the pronounced bias towards a later phase in the detected systoles, we consider the automatically determined end-systolic phases on equal terms to the manually given.

Quantitatively, we observe that our method handles challenges such as epicardial fat well due to its constrained nature. See e.g. Figure 13.7. However, inspection of the obtained results revealed that determination of the extreme basal and apical slice still poses a challenge as it does for the manual observer. This is primarily due to the anisotropy of the voxels, which only allows for a slice-based annotation. To avoid this, high-definition models can be built based on data from sophisticated acquisition schemes such as [29] where a much improved coverage of the LV is

obtained. Such models could subsequently be used to search sparse, short-axis slice-oriented data similar to the data in this study. We anticipate that results from such a strategy will outperform human performance due to the vastly more detailed prior model.

The obtained volume errors were fair except for one outlier in both size and shape that also showed pronounced image artefacts (patient 5). Epicardial volumes (used for calculating the mass of myocardium) showed a relative error, which was markedly higher than for the endocardial volumes.

Although not directly comparable due to differences in scanners, pulse sequences, subjects, et cetera, the variability of the obtained ejection fraction estimates compares very well to the inter-observer variability in a recently published MRI-based ejection fraction study [177]. Further, limits of agreement for end-systolic volumes and end-diastolic volumes are near-identical to those presented in [169], which concluded on basis of these that a new method for assessing ESV and EDV was accurate and produced results comparable to the established method. Due to this similarity, we conclude that there is a clear indication that the presented method can replace the current resource-demanding and tedious practice of manual annotation required to estimate cardiac blood pool volumes and ejection fraction.

The usage of texture model pruning, whiskers and decoupling of shape and texture eigenspaces in combination was showed to have a markedly positive influence on the results. So did the correction of respiratory induced slice displacements. However, using full resolution images compared to a decimation in the x and y axes increased the error slightly. This suggests that the regularisation obtained by the subsampling was desirable, i.e. it eliminated more noise structures, than it deteriorated the latent tissue structure. In this conjunction, it may be interesting to apply a more principled and data-driven regularisation using a truncated wavelet texture representation as introduced for AAMs in [250] and further explored in [218].

Since this work (arguably) represents the first application of 3D AAMs independent from the work by Mitchell and co-workers [161] some comparisons may be of value to the reader to outline differences and extensions. Contrary to the single image 3D AAM in [161], this work features a bi-temporal 3D AAM applied to 4D images using an automated systole detection. The tetrahedral warper employs a general 3D Delaunay decomposition, contrary to application-tailored tetrahedral meshes. Ordinary Procrustes alignment was based on a simple SVD-based solution, contrary to – admittedly elegant, but not so well-known and more involved – quaternion-based alignment. The problem of respiratory motion mentioned in [161] has been addressed and assessed in our data. Non-Gaussian texture appearance caused by papillary muscles has been addressed by a texture model pruning mechanism. A simple method of adding proximity sampling to the texture model was employed, contrary to application-tailored tetrahedra. Running multiple searches, contrary to a dedicated search method initialised our models. Finally, our

timings indicate a substantially faster implementation. In [161], it took 2–3 minutes to optimise a single-image 3D AAM containing approximately 7000 voxels using a 1 GHz PC. In our setting, a 3×3 grid search using a bi-temporal 3D AAM containing approximately 22000 voxels took 3.4 seconds on average using a 2.4 GHz PC. Although this indicates a significant difference in performance, we emphasise that a direct comparison is not possible, due to the different requirements stemming from different data sets, et cetera.

We hypothesize three major issues that limit the accuracy obtainable from the described approach. Primarily, the shape PCA may not be able to approximate new examples from a training set of only eleven examples. This may also be true for the texture model, which may cause uninteresting minima in the cost function evaluated during the model-to-image fitting process. Judged from the image appearance in the data set, we believe the shape PCA to be the limiting factor. Next is the issue of the extreme basal and apical slice, which is more a data problem than a model problem. As this also pertains to the ground truth solution it may be hard to address and evaluate. In this matter, we have outlined a solution above based on additional image data. Finally, the naïve approach taken to establishment of point correspondences may prove insufficient. In particular, if more detailed models of the myocardium, or local measures such as wall thickness, et cetera, are required in future studies. Fortunately, this topic is well treated in the literature and some alternatives are given above.

As a concluding remark; statistical approaches similar to the described method are ultimately limited by the sample size and how representative the training set is. Consequently, hard conclusions regarding the general performance of such a method will remain fragile until more training data becomes available.

13.7 Conclusion

A method for unsupervised estimation of ventricular volumes, masses and ejection fraction from four-dimensional cardiac cine MRI has been presented. The learning-based method establishes statistical models of myocardial shape and appearance from a training set of end-diastolic and end-systolic images, which enables it to rapidly infer global cardiac indices from unseen 4D images contaminated with substantial image artefacts.

Cross-validation of the method carried out on clinical-quality scans of twelve volunteers indicates that ejection fraction and cardiac blood pool volumes can be estimated automatically and rapidly with an accuracy comparable to the typical inter-observer variability.

Acknowledgements

M.D. Dorthe Pedersen, Danish Research Centre for Magnetic Resonance, DRCMR, Hvidovre University Hospital is gratefully acknowledged for providing MR data and ground truth two-dimensional myocardial contours for this study. The work of M. B. Stegmann was supported by the Danish Medical Research Council, grant no. 52-00-0767.

13.A Ordinary Procrustes Alignment in k Dimensions

This appendix serves to demonstrate the validity of the solution to the orthogonal ordinary Procrustes alignment problem for two shapes in k dimensions (in our case $k = 3$). Let $\mathbf{X}_1$, $\mathbf{X}_2 \in \mathbb{R}^{N \times k}$ denote two shapes with centroids incident with the origin. The optimal rotation – in a least-squares sense – that superimposes $\mathbf{X}_2$ onto $\mathbf{X}_1$, is given by

$$\arg\min_{\mathbf{R}} ||\mathbf{X}_1 - \mathbf{X}_2\mathbf{R}||^2, \tag{13.15}$$

where $||\mathbf{M}|| = \sqrt{\mathrm{tr}(\mathbf{M}^\mathsf{T}\mathbf{M})}$ denotes the Frobenius norm, and $\mathbf{R}$ is an orthonormal matrix subject to $\det(\mathbf{R}) = 1$. Now consider the following expansion

$$\begin{aligned} \mathbf{P} &= (\mathbf{X}_1 - \mathbf{X}_2\mathbf{R})^\mathsf{T}(\mathbf{X}_1 - \mathbf{X}_2\mathbf{R}) \\ &= \mathbf{X}_1^\mathsf{T}\mathbf{X}_1 - \mathbf{X}_1^\mathsf{T}\mathbf{X}_2\mathbf{R} - \mathbf{R}^\mathsf{T}\mathbf{X}_2^\mathsf{T}\mathbf{X}_1 + \mathbf{R}^\mathsf{T}\mathbf{X}_2^\mathsf{T}\mathbf{X}_2\mathbf{R}, \end{aligned} \tag{13.16}$$

and its trace

$$\begin{aligned} \mathrm{tr}(\mathbf{P}) &= \mathrm{tr}(\mathbf{X}_1^\mathsf{T}\mathbf{X}_1) - \mathrm{tr}(\mathbf{X}_1^\mathsf{T}\mathbf{X}_2\mathbf{R}) - \mathrm{tr}((\mathbf{R}^\mathsf{T}\mathbf{X}_2^\mathsf{T}\mathbf{X}_1)^\mathsf{T}) + \mathrm{tr}((\mathbf{R}^\mathsf{T}\mathbf{X}_2^\mathsf{T}\mathbf{X}_2\mathbf{R})^\mathsf{T}) \\ &= \mathrm{tr}(\mathbf{X}_1^\mathsf{T}\mathbf{X}_1 + \mathbf{X}_2^\mathsf{T}\mathbf{X}_2) - 2\,\mathrm{tr}(\mathbf{X}_1^\mathsf{T}\mathbf{X}_2\mathbf{R}). \end{aligned} \tag{13.17}$$

Minimising (13.15) is thus equivalent to maximising $\mathrm{tr}(\mathbf{X}_1^\mathsf{T}\mathbf{X}_2\mathbf{R})$. This is achieved if, and only if, $\mathbf{X}_1^\mathsf{T}\mathbf{X}_2\mathbf{R}$ is symmetric and positive semi definite. See [231] for a proof.

Consider the following singular value decomposition (see e.g. [106, 180])

$$\mathbf{X}_1^\mathsf{T}\mathbf{X}_2 = \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}, \tag{13.18}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices (called singular vectors) and $\mathbf{S}$ is a diagonal matrix of singular values. The optimal orthonormal solution to (13.15) is claimed to be

$$\mathbf{R} = \mathbf{V}\mathbf{U}^\mathsf{T}. \tag{13.19}$$

It now follows directly that

$$\mathbf{X}_1^\mathsf{T}\mathbf{X}_2\mathbf{R} = \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}\mathbf{V}\mathbf{U}^\mathsf{T} = \mathbf{U}\mathbf{S}\mathbf{U}^\mathsf{T} \tag{13.20}$$

is symmetric and positive semi definite since all eigenvalues – given by $\mathbf{S}$ – are nonnegative.

However, $\mathbf{V}\mathbf{U}^\mathsf{T}$ only holds a rotation matrix when $\mathbf{U}\mathbf{S}\mathbf{U}^\mathsf{T}$ is symmetric and positive definite. In the rare case where not all singular values are nonnegative, the correlation matrix, $\mathbf{X}_1^\mathsf{T}\mathbf{X}_2\mathbf{R}$, can be maximized using a reflection, i.e. $\det(\mathbf{R}) = -1$. A solution to (13.15) that ensures rotation is

$$\mathbf{R} = \mathbf{V}\mathbf{\Delta}\mathbf{U}^\mathsf{T}, \tag{13.21}$$

where $\mathbf{\Delta} = \mathrm{diag}(\,[\,1\ \ldots\ 1\ \det(\mathbf{V}\mathbf{U}^\mathsf{T})\,]\,) \in \mathbb{R}^{k \times k}$. Refer to [7, 80, 118, 120, 231, 240] for further elaborations.

13.B Barycentric Coordinates of a Tetrahedron

Let $\mathbf{b} = [\ \alpha\ \beta\ \gamma\ \delta\ ]^{\mathsf{T}}$ denote the barycentric coordinates of a simplex in three dimensions; a tetrahedron. Subject to the two constraints, $1 = \alpha + \beta + \gamma + \delta$ and $0 \leq \alpha, \beta, \gamma, \delta \leq 1$, these coordinates will uniquely determine any position, $\mathbf{p}$, within the tetrahedron, $\mathbf{T}$ (given by the point set $\{x_i, y_i, z_i\}_1^4$),

$$\mathbf{p} = \mathbf{T}\mathbf{b} \quad \Leftrightarrow \quad \begin{bmatrix} 1 \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix}. \tag{13.22}$$

Provided that $\mathbf{T}$ has full rank the barycentric coordinates corresponding to the point $\mathbf{p}$ are thus given by $\mathbf{b} = \mathbf{T}^{-1}\mathbf{p}$ and expanded below, where $\det(\mathbf{T})$ is denoted by d

$$\begin{aligned} \alpha \ = -\tfrac{1}{d}(\ & -x_3y_2z + x_3yz_2 + yx_2z_4 - z_4xy_2 + z_4xy_3 - y_4x_2z - \\ & yx_4z_2 + y_2zx_4 - x_2z_3y + zx_3y_4 - zx_4y_3 - z_4x_3y + \\ & z_3yx_4 - z_3y_4x + xy_2z_3 - xy_3z_2 + x_2zy_3 + y_4xz_2 - \\ & x_4y_2z_3 + x_4y_3z_2 + x_3y_2z_4 - x_3y_4z_2 + x_2z_3y_4 - x_2z_4y_3 \ ), \end{aligned} \tag{13.23}$$

$$\begin{aligned} \beta \ = \tfrac{1}{d}(\ & -z_1yx_4 - z_1xy_3 + z_1y_4x - z_1x_3y_4 + z_1x_4y_3 + z_1x_3y - \\ & x_1z_3y - x_3y_1z + z_4xy_3 + yx_1z_4 + y_1zx_4 - y_1z_4x - \\ & y_4x_1z + zx_3y_4 - zx_4y_3 - z_4x_3y + z_3yx_4 - z_3y_4x + \\ & xy_1z_3 + x_1zy_3 - x_1z_4y_3 + x_1z_3y_4 - x_4y_1z_3 + x_3y_1z_4 \ ), \end{aligned} \tag{13.24}$$

$$\begin{aligned} \gamma \ = -\tfrac{1}{d}(\ & -x_2y_4z_1 + x_4z_1y_2 - xz_1y_2 + x_2yz_1 - z_1yx_4 + z_1y_4x + \\ & xz_2y_1 - y_4x_1z - y_1z_4x - y_2zx_4 - x_1yz_2 + x_1y_4z_2 - \\ & x_4z_2y_1 + x_1zy_2 - y_4xz_2 - x_1z_4y_2 - yx_2z_4 + yx_4z_2 + \\ & yx_1z_4 + z_4xy_2 + x_2y_1z_4 - x_2y_1z + y_4x_2z + y_1zx_4 \ ), \end{aligned} \tag{13.25}$$

$$\begin{aligned} \delta \ = \tfrac{1}{d}(\ & -xz_1y_2 + z_1xy_3 - x_2y_3z_1 + x_2yz_1 - z_1x_3y + x_3z_1y_2 - \\ & x_2y_1z + x_1z_3y - x_3y_2z + x_3yz_2 + x_3y_1z - x_1z_3y_2 - \\ & x_2z_3y - x_1yz_2 + xz_2y_1 + x_1y_3z_2 - x_3z_2y_1 - xy_1z_3 + \\ & xy_2z_3 - xy_3z_2 + x_1zy_2 - x_1zy_3 + x_2zy_3 + x_2y_1z_3 \ ), \end{aligned} \tag{13.26}$$

$$\begin{aligned} d \ = \ & -x_2y_3z_1 + x_2y_4z_1 - z_1x_3y_4 - x_4z_1y_2 + z_1x_4y_3 + \\ & x_3z_1y_2 - x_1z_3y_2 + x_1y_3z_2 - x_3z_2y_1 + x_2y_1z_3 - x_1y_4z_2 + \\ & x_4z_2y_1 - x_2y_1z_4 - x_1z_4y_3 + x_4y_2z_3 + x_1z_3y_4 - x_4y_3z_2 - \\ & x_4y_1z_3 + x_3y_1z_4 - x_3y_2z_4 + x_3y_4z_2 - x_2z_3y_4 + x_2z_4y_3 + x_1z_4y_2. \end{aligned} \tag{13.27}$$

Degenerate tetrahedra (having coplanar points) will thus have $d = 0$, i.e. no inverse.

CHAPTER 14

Appendix

This chapter briefly describes three software components developed during the thesis period. All are publicly available either in source code form and/or in binary form on the internet.

These components can also be found on the enclosed CD along with a hyperlinked edition of this thesis in the Adobe Portable Document Format (PDF).

Finally, the last section describes how a basis can be constructed that decorrelates shape variation.

14.A The AAM-API

The AAM-API is a C++ implementation of the Active Appearance Model framework. Emphasis is put on performance to yield responsiveness for interactive applications, while yet maintaining a clear structure. The API is focused on single-image 2D applications, but the modularity enables extensions, such as coupled 2D time series and fusion of 2D images from different modalities, e.g. PET, CT and MRI. Such extensions have previously been described in the literature.

To facilitate transparency in the model building process a vast range of documentation can be exported. This makes the package well suited for educational purposes. Further, a method for model initialisation is included, thus rendering image segmentation and analysis a fully automated process. Example segmentations obtained using the API are shown in Figure 14.1.

Alongside the API are scripts for annotating images etc. and the binary AAM-Lab; a graphical user interface for manipulating annotations and running model searches. The latter is very useful for building new models bootstrapped from a small set of annotated examples. On OpenGL-compliant machines the AAMLab also features an exploratory real-time visualisation of the AAM deformation modes.

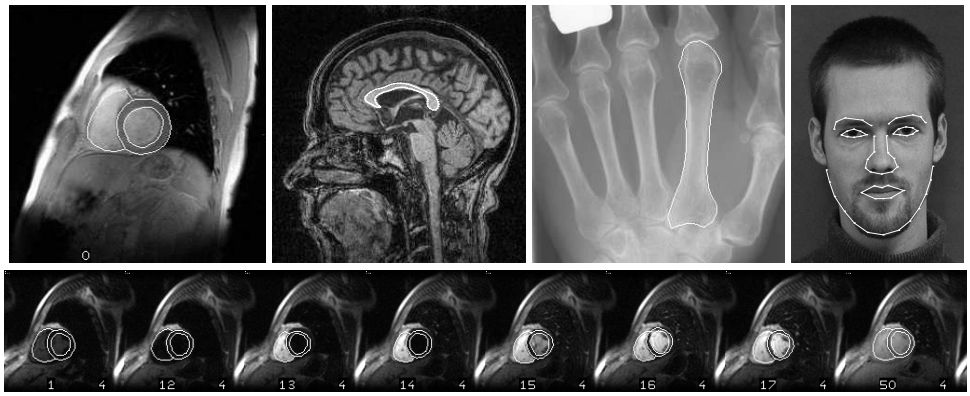


Figure 14.1: Examples of the medical applications of the AAM-API. From top, left: Cardiac cine MRI, brain MRI, metacarpal radiograph, face image and cardiac perfusion MRI.

A more thorough introduction to the package and accompanying reference data sets is given in [213]. Finally, several performance benchmarks can be generated in e.g. a leave-one-out study as a part of a clinical validation.

The software runs on the Microsoft Windows platform and requires two 3rd part libraries, namely LAPACK and VisionSDK, both available in source code form. Microsoft Visual Studio is required to modify or recompile the software. The complete package is described in more detail and is available for download from

http://www.imm.dtu.dk/∼aam/.

Source code documentation and papers describing the applications shown in Figure 14.1 are also available from this address.

14.B The 4D Cardiac Viewer

The 4D Cardiac Viewer is a tool for interactively investigating dynamic volumetric images of the human heart. Its primary goal is to aid anatomical understanding and allow inspection of various dynamic behaviours. Further, it allows for loading of spatio-temporal mesh data, which is superimposed on the volume data. This program provided all figures showing cardiac MRI data in Chapter 12 and Chapter 13.

An example 4D cardiac cine image is included (cardiacvol.tvi) with an associated mesh file (cardiacvol.obj) containing annotated endocardial surfaces at end-diastole and end-systole in addition to an end-systolic epicardial surface.

Format descriptions for the TVI and OBJ formats are also included along with matlab scripts for TVI and OBJ input/output. Figure 14.2 shows the user interface of the 4D Cardiac viewer.

Refer to http://www.imm.dtu.dk/~aam/ for further information.

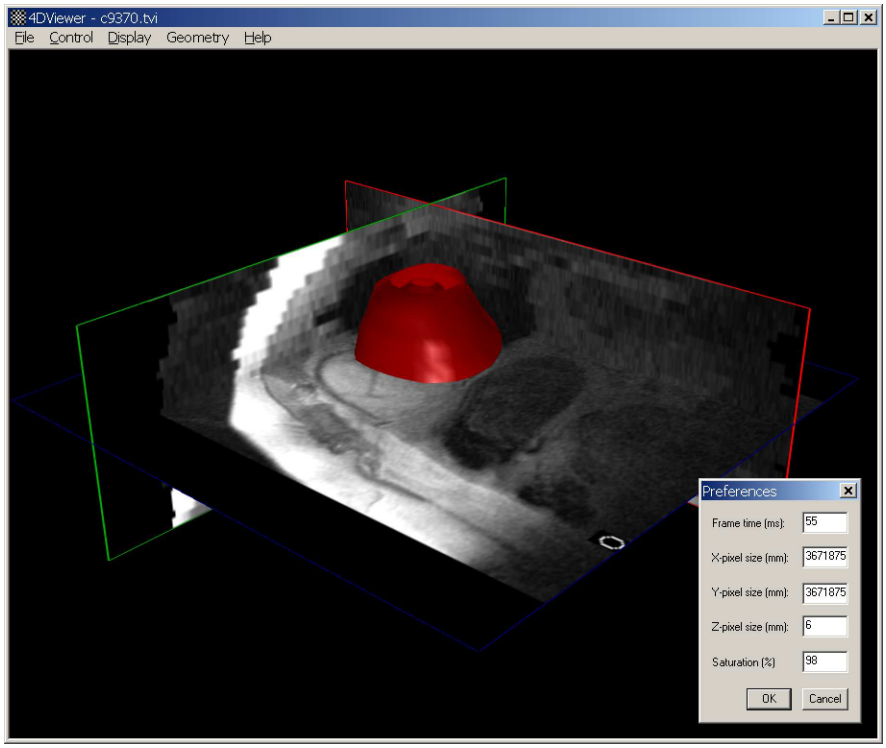


Figure 14.2: The user interface of the 4D Cardiac viewer application.

14.C The AAM Explorer

The AAM Explorer is a tool for interactively investigating the modes of an Active Appearance Model. In this case an AAM built on 35 frontal face images.

The implementation was done using the AAM-API and OpenGL for rapid visualization. In the current version a 30000-pixel, 27-parameter AAM is deformed in real-time (50 Hz on an Athlon 1.2GHz using the Intel Math Kernel library for matrix computations). Though having 27 parameters, the interface only shows six sliders for clarity, since the program was originally made for a Danish exhibition targeted at a broad audience. The first six modes account for 56% of the total variation in the training set. Refer to http://www.imm.dtu.dk/~aam/ for further information.

Minimum System Requirements

- Microsoft Windows 2000 / XP (or newer)
- An OpenGL accelerated graphics board (with properly installed OpenGL drivers)
- 800 MHz Processor / 64 MB RAM

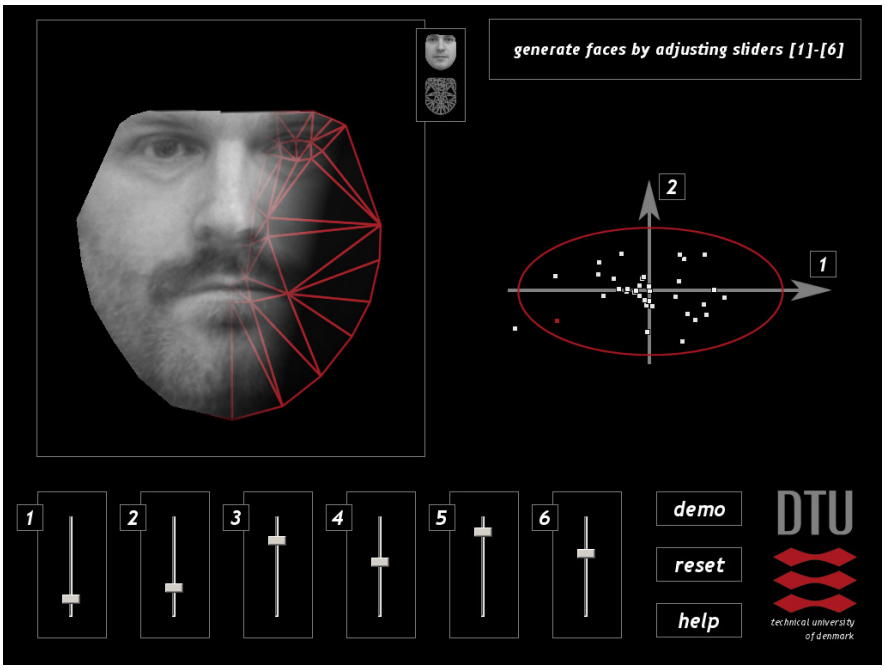


Figure 14.3: The user interface of the AAM Explorer application.

14.D Shape Decomposition by PCA

This appendix contains a self-contained presentation of how the Principal Component Analysis (PCA) can be derived by means of simple linear algebra.

Consider the case of having N planar shapes consisting of n points, where each shape is represented as:

$$\mathbf{x} = [\, x_1 \; x_2 \; \ldots \; x_n \; y_1 \; y_2 \; \ldots \; y_n \,]^\mathsf{T} \tag{14.1}$$

Looking at the covariance and correlation matrix of such shapes we typically observe a pattern similar to the one shown in Figure 14.4.

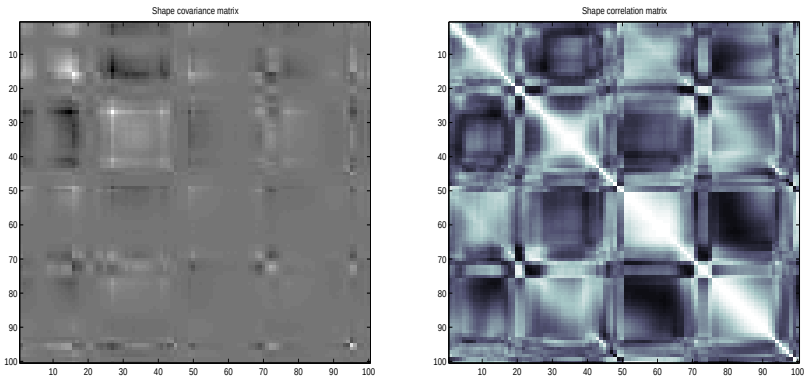


Figure 14.4: Shape covariance matrix, white=positive, black=negative, grey=zero (left). Shape correlation matrix (right).

This suggest that there might exist a shape representation accounting for the obvious correlation between points. If some point movements were to be (numerically) highly correlated, this could be exploited to reduce dimensionality. In this case, we will seek a linear transformation of our data:

$$\mathbf{y} = \mathbf{M}\mathbf{x} \tag{14.2}$$

First, consider the mean shape $\overline{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$ and the estimate of the shape covariance matrix:

$$\mathbf{\Sigma}_\mathbf{x} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^\mathsf{T} \tag{14.3}$$

The mean of the $\mathbf{y}$-variables can then be expressed as:

$$\overline{\mathbf{y}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i = \frac{1}{N}\sum_{i=1}^{N}\mathbf{M}\mathbf{x}_i = \mathbf{M}\overline{\mathbf{x}} \tag{14.4}$$

And consequently the estimate of the covariance of the $\mathbf{y}$'s:

$$\begin{aligned}
\mathbf{\Sigma}_\mathbf{y} &= \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{y}_i - \overline{\mathbf{y}})(\mathbf{y}_i - \overline{\mathbf{y}})^\mathsf{T} \\
&= \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{M}\mathbf{x}_i - \mathbf{M}\overline{\mathbf{x}})(\mathbf{M}\mathbf{x}_i - \mathbf{M}\overline{\mathbf{x}})^\mathsf{T} \\
&= \frac{1}{N-1}\sum_{i=1}^{N}\mathbf{M}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{M}(\mathbf{x}_i - \overline{\mathbf{x}}))^\mathsf{T} \\
&= \frac{1}{N-1}\sum_{i=1}^{N}\mathbf{M}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^\mathsf{T}\mathbf{M}^\mathsf{T} \\
&= \mathbf{M}\left(\frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^\mathsf{T}\right)\mathbf{M}^\mathsf{T} \\
&= \mathbf{M}\mathbf{\Sigma}_\mathbf{x}\mathbf{M}^\mathsf{T}
\end{aligned} \tag{14.5}$$

Then, if we limit ourselves to orthogonal transformations (i.e. $\mathbf{M}^{-1} = \mathbf{M}^\mathsf{T}$) left-multiplication by $\mathbf{M}^\mathsf{T}$ in (14.5) yields:

$$\mathbf{M}^\mathsf{T}\mathbf{\Sigma}_\mathbf{y} = \mathbf{\Sigma}_\mathbf{x}\mathbf{M}^\mathsf{T} \tag{14.6}$$

Substitution of $\mathbf{M}^\mathsf{T}$ by $\mathbf{\Phi}$ yields:

$$\mathbf{\Sigma}_\mathbf{x}\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Sigma}_\mathbf{y} \tag{14.7}$$

From (14.7) it is seen that if $\mathbf{\Phi}$ is chosen as the (column) eigenvectors of the symmetric matrix $\mathbf{\Sigma}_\mathbf{x}$, then the covariance of the transformed shapes, $\mathbf{\Sigma}_\mathbf{y}$, becomes a diagonal matrix of eigenvalues. In the case of correlated points the smallest eigenvalues will be (close to) zero and the corresponding eigenvectors could be omitted from $\mathbf{\Phi}$, thus reducing the length of $\mathbf{y}$.

In conclusion, to establish an orthogonal linear transform that decorrelates data vectors, the transformation matrix must be the eigenvectors of the covariance matrix of the original data. In order to back transform from the new set of variables, $\mathbf{y}$, we invert (14.2), remembering that $\mathbf{M}$ is orthogonal:

$$\mathbf{x} = \mathbf{M}^{-1}\mathbf{y} = \mathbf{M}^\mathsf{T}\mathbf{y} = \mathbf{\Phi}\mathbf{y} \tag{14.8}$$

As a final comment; one would typically (as in the AAM case) apply PCA on variables with zero mean:

$$\mathbf{y} = \mathbf{M}(\mathbf{x} - \overline{\mathbf{x}}) \quad , \quad \mathbf{x} = \overline{\mathbf{x}} + \mathbf{\Phi}\mathbf{y} \tag{14.9}$$

This method of dealing with redundancy in multivariate data is known as *Principal Component Analysis* (PCA) or the *Karhunen-Loéve Transform* (KLT). Refer to a standard textbook on multivariate statistics for further details.

List of Figures

List of Tables

List of Algorithms

Bibliography

[1] N. Ablitt, J. Gao, P. Gatehouse, and G. Z. Yang. Motion decoupling and registration for 3D magnetic resonance myocardial perfusion imaging. In *International Conference Computational Science - ICCS 2002*, volume 2331, pages 285–294. Springer-Verlag, 2002.

[2] P. T. Adeler. *Hemodynamic simulation of the heart using a 2D model and MR data*. PhD thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, 2001.

[3] J. Ahlberg. *Model-based Coding – Extraction, Coding and Evaluation of Face Model Parameters*. PhD thesis, Department of Electrical Engineering, Linköping University, Sweden, 2002.

[4] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, third edition, 1999. ISBN 0-89871-447-8 (paperback). Web page: http://www.netlib.org/lapack/.

[5] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. *Image Processing, IEEE Transactions on*, 1(2):205–220, 1992. ISSN 10577149.

[6] S. G. Armato, M. L. Giger, and H. MacMahon. Automated lung segmentation in digitized posterior-anterior chest radiographs. *Academic Radiology*, 4:245–255, 1998.

[7] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Trans. on Pattern Anal. and Machine Intelligence*, 9(5):698–700, 1987. ISSN 01628828.

[8] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu. An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. *Journal of the ACM*, 45(6):891–923, 1998.

[9] M. A. Audette, F. P. Ferrie, and T. M. Peters. An algorithmic overview of surface registration techniques for medical imaging. *Medical Image Analysis*, 4(3):201–217, 2000. ISSN 13618415.

[10] S. Baker, R. Gross, and I. Matthews. Lucas-Kanade 20 years on: A unifying framework: Part 3. Technical Report CMU-RI-TR-03-35, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, November 2003.

[11] S. Baker, R. Gross, I. Matthews, and T. Ishikawa. Lucas-Kanade 20 years on: A unifying framework: Part 2. Technical Report CMU-RI-TR-03-01, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, February 2003.

[12] S. Baker and I. Matthews. Equivalence and efficiency of image alignment algorithms. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society*

Conference on, volume 1, pages I–1090. IEEE Comput. Soc, 2001. ISBN 0769512720.

[13] S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework: Part 1. Technical Report CMU-RI-TR-02-16, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, July 2002.

[14] S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 2004.

[15] R. Bansal and G. Funka-Lea. Integrated image registration for cardiac MR perfusion data. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2002. 5th International Conference. Proceedings, Part I (Lecture Notes in Computer Science Vol.2488)*, pages 659–66, 2002.

[16] H. J. Barnhard, J. A. Pierce, J. W. Joyce, and J. H. Bates. Roentgenographic determination of total lung capacity. A new method evaluated in health, emphysema and congestive heart failure. *American Journal of Medicine*, 28:51–60, 1960.

[17] A. U. Batur and M. H. Hayes. A novel convergence scheme for active appearance models. *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, 1:359–366, 2003. ISSN 10636919.

[18] F. Behloul, J. P. Roux, and R. Unterreiner. Automatic boundary detection of the left ventricle and measurement of local myocardial perfusion in MRI. *Computers in Cardiology 1997*, pages 145–148, 1997.

[19] R. Beichel, G. Gotschuli, E. Sorantin, F. W. Leberl, and M. Sonka. Diaphragm dome surface segmentation in CT data sets: a 3D active appearance model approach. *SPIE Medical Imaging 2002*, pages 475–484, 2002.

[20] R. Beichel, S. Mitchell, E. Sorantin, F. Leberl, A. Goshtasby, and M. Sonka. Shape- and appearance-based segmentation of volumetric medical images. *IEEE International Conference on Image Processing*, 2:589–592, 2001.

[21] F. Bellemare, J. Couture, M. Cordeau, P. Leblanc, and E. Lafontaine. Anatomic landmarks to estimate the length of the diaphragm from chest radiographs: effects of emphysema and lung volume reduction surgery. *Chest*, 120(2):444–452, 2001.

[22] J. C. Bezdek, L. O. Hall, and L. P. Clarke. Review of MR image segmentation techniques using pattern recognition. *Medical Physics*, 20:1033–1048, 1993.

[23] L. M. Bidaut and J. P. Vallee. Automated registration of dynamic MR images for the quantification of myocardial perfusion. *Jour. Magn. Reson. Imaging*, 13(4):648–655, 2001.

[24] Michael J. Black and Allan D. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. *International Journal of Computer Vision*, 26(1):63–84, 1998. ISSN 09205691.

[25] A. Blake and M. Isard. *Active Contours*. Springer, 1998.

[26] J. M. Bland and D. G. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 1(8476):307–310, 1986.

[27] J. M. Bland and D. G. Altman. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet*, 346(8982):1085–1087, 1995.

[28] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. *Computer Graphics Proceedings. SIGGRAPH 99*, pages 187–94, 1999.

[29] D. C. Bloomgarden, Z. A. Fayad, V. A. Ferrari, B. Chin, M. G. St. John Sutton, and L. Axel. Global cardiac function using fast breath-hold mri: Validation of new acquisition and analysis techniques. *Magnetic Resonance in Medicine*, 37(5):683–92, 1997. ISSN 07403194.

[30] F. L. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–85, 1989.

[31] F. L. Bookstein. Landmark methods for forms without landmarks: localizing group differences in

outline shape. *Medical Image Analysis*, 1(3):225–244, 1997.

[32] A. Bornstedt, E. Nagel, S. Schalla, B. Schnackenburg, C. Klein, and E. Fleck. Multi-slice dynamic imaging: Complete functional cardiac MR examination within 15 seconds. *Journal of Magnetic Resonance Imaging*, 14(3):300–305, 2001.

[33] H. G. Bosch, S. C. Mitchell, B. P. F. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J. H. C. Reiber. Active appearance-motion models for endocardial contour detection in time sequences of echocardiograms. *Proceedings of SPIE - The International Society for Optical Engineering*, 4322(1):257–268, 2001. ISSN 0277786x.

[34] H. G. Bosch, S. C. Mitchell, B. P. F. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J. H. C. Reiber. Active appearance-motion models for fully automated endocardial contour detection in time sequences of echocardiograms. *International Congress Series*, 1230:941–947, 2001. ISSN 05315131.

[35] J. G. Bosch, S. C. Mitchell, B. P. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J. H. Reiber. Fully automated endocardial contour detection in time sequences of echocardiograms by three-dimensional active appearance models. *SPIE Medical Imaging 2002*, pages 452–462, 2002.

[36] J. G. Bosch, S. C. Mitchell, B. P. F. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J. H. C. Reiber. Fully automated endocardial contour detection in time sequences of echocardiograms by active appearance motion models. *Computers in Cardiology*, pages 93–96, 2001. ISSN 02766574.

[37] J. G. Bosch, S. C. Mitchell, B. P. F. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J. H. C. Reiber. Automatic segmentation of echocardiographic sequences by active appearance motion models. *Medical Imaging, IEEE Transactions on*, 21(11):1374–1383, 2002. ISSN 02780062.

[38] R. Bowden. *Learning non-linear Models of Shape and Motion*. PhD thesis, Dept Systems Engineering, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK., 1999.

[39] K. W. Bowyer, M. H. Loew, H. S. Stiehl, and M. A. Viergever. Methodology of evaluation in medical image computing. In *Rep. Dagstuhl Workshop*, 2001. URL http://www.dagstuhl.de/01111/Report.

[40] L. Bracoud, F. Vincent, C. Pachai, E. Canet, P. Croisille, and D. Revel. Automatic registration of MR first-pass myocardial perfusion images. In *Functional Imaging and Modeling of the Heart, FIMH 2003*, volume 2674 of *Lecture Notes in Computer Science*, pages 215 – 223, Lyon, France, 2003. Springer Verlag.

[41] M. Breeuwer, M. Quist, L. Spreeuwers, I. Paetsch, N. Al-Saadi, and E. Nagel. Towards automatic quantitative analysis of cardiac MR perfusion images. *International Congress Series*, 1230:967–973, 2001. ISSN 05315131.

[42] M. Breeuwer, L. Spreeuwers, and M. Quist. Automatic quantitative analysis of cardiac MR perfusion images. *Proceedings of the SPIE - The International Society for Optical Engineering*, 4322(1-3): 733–42, 2001. ISSN 0277786x.

[43] M. Brejl and M. Sonka. Object localization and border detection criteria design in edge-based image segmentation: automated learning from examples. *Medical Imaging, IEEE Transactions on*, 19 (10):973–985, 2000. ISSN 02780062.

[44] M. S. Brown, L. S. Wilson, B. D. Doust, R. W. Gill, and C. Sun. Knowledge-based method for segmentation and analysis of lung boundaries in chest X-ray images. *Computerized Medical Imaging and Graphics*, 22:463–477, 1998.

[45] P. J. Burt and E. H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, COM-31,4:532–540, 1983.

[46] M. D. Cerqueira, N. J. Weissman, V. Dilsizian, A. K. Jacobs, S. Kaul, W. K. Laskey, D. J. Pennell, J. A. Rumberger, T. Ryan, and M. S. Verani. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: A statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the american heart association. *Circulation - Hagertown*, 105(4):539, 2002. ISSN 00097322.

[47] C. A. Cocosco, A. P. Zijdenbos, and A. C. Evans. A fully automatic and robust brain MRI tissue classification method. *Medical Image Analysis*, 7(4):513–527, 2003.

[48] T. F. Cootes, C. Beeston, G. J. Edwards, and C. J. Taylor. A unified framework for atlas matching using active appearance models. In *Information Processing in Medical Imaging. IPMI'99*, pages 322–33. Springer-Verlag, 1999.

[49] T. F. Cootes, D. Cooper, C. J. Taylor, and J. Graham. A trainable method of parametric shape description. In *Proc. British Machine Vision Conference*, pages 54–61. Springer-Verlag, 1991.

[50] T. F. Cootes, D. H. Cooper, C. J. Taylor, and J. Graham. Trainable method of parametric shape description. *Image and Vision Computing*, 10(5):289–94, 1992. ISSN 02628856.

[51] T. F. Cootes, G. Edwards, and C. J. Taylor. A comparative evaluation of active appearance model algorithms. In *BMVC 98. Proc. of the Ninth British Machine Vision Conf.*, volume 2, pages 680–689. Univ. Southampton, 1998.

[52] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proc. European Conf. on Computer Vision*, volume 2, pages 484–498. Springer, 1998.

[53] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Comparing active shape models with active appearance models. In *Proc. British Machine Vision Conf.*, pages 173–182, 1999.

[54] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[55] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam. The use of active shape models for locating structures in medical images. *Image and Vision Computing*, 12(6):355–366, 1994.

[56] T. F Cootes and P. Kittipanya-ngam. Comparing variations on the active appearance model algorithm. In *Proceedings of the British Machine Vision Conference, BMVC*, volume 2, pages 837–846, 2002.

[57] T. F. Cootes and Taylor. Active shape models – 'smart snakes'. In *Proc. British Machine Vision Conf., BMVC92*, pages 266–275, 1992.

[58] T. F. Cootes and C. J. Taylor. Combining point distribution models with shape models based on finite element analysis. In *Proc. British Machine Vision Conference, BMVC*, volume 1, pages 419–428, 1994.

[59] T. F. Cootes and C. J. Taylor. Combining point distribution models with shape models based on finite element analysis. *Image and Vision Computing*, 13(5):403–9, 1995.

[60] T. F. Cootes and C. J. Taylor. A mixture model for representing shape variation. In *Proceedings of the 8th British Machine Vision Conference, BMVC*, volume 1, pages 110–119. BMVA Press, 1997.

[61] T. F. Cootes and C. J. Taylor. A mixture model for representing shape variation. *Image and Vision Computing*, 17(8):567–574, 1999.

[62] T. F. Cootes and C. J. Taylor. Combining elastic and statistical models of appearance variation. In *Proc. European Conf. on Computer Vision*, volume 1, pages 149–163, 2000.

[63] T. F. Cootes and C. J. Taylor. Constrained active appearance models. *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 1:748–754 vol.1, 2001.

[64] T. F. Cootes and C. J. Taylor. On representing edge structure for model matching. In *Proc. IEEE Computer Vision and Pattern Recognition – CVPR*, volume 1, pages 1114–1119. IEEE, 2001.

[65] T. F. Cootes and C. J. Taylor. *Statistical Models of Appearance for Computer Vision*. Tech. report, University of Manchester, 2001. URL http://www.isbe.man.ac.uk/~bim/.

[66] T. F. Cootes and C. J. Taylor. Statistical models of appearance for medical image analysis and computer vision. In *Proc. SPIE Medical Imaging 2001*, volume 1, pages 236–248. SPIE, 2001.

[67] T. F. Cootes, C. J. Taylor, D. Cooper, and J. Graham. Training models of shape from sets of examples. In *Proc. British Machine Vision Conf., BMVC92*, pages 9–18, 1992.

[68] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models – their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[69] T. F. Cootes, C. J. Taylor, C. Twining, and N. Thacker. A framework for building deformable atlases.

Technical report, University of Manchester, 2002. URL www.isbe.man.ac.uk/~bim/Papers/def_atlas.ps.

[70] T. F. Cootes, K. Walker, and C. J. Taylor. View-based active appearance models. In *Proc. 4th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 227–32. IEEE Comput. Soc, 2000.

[71] T. F. Cootes, G. V. Wheeler, K. N. Walker, and C. J. Taylor. Coupled-view active appearance models. *BMVC 2000*, pages 52–61 vol.1, 2000.

[72] M. G. Danilouchkine, J. J. M. Westenberg, H. J. Lamb, J. H. C. Reiber, and B. P. F. Lelieveldt. Accuracy of fully automatic vs. manual planning of cardiac MR acquisitions. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003, 6th Int. Conference, Montréal, Canada*, LNCS 2879, pages 961–962. Springer, 2003.

[73] S. Darkner, R. Larsen, M. B. Stegmann, and B. K. Ersbøll. Wedgelet enhanced appearance models. In *2nd International Workshop on Generative Model-Based Vision – GMBV, CVPR 2004*, 2004 (to appear).

[74] Image Sciences Institute Research Databases. URL http://www.isi.uu.nl/Research/Databases/.

[75] R. H. Davies. *Learning Shape: Optimal Models for Analysing Natural Variability*. PhD thesis, Department of Imaging Science and Biomedical Engineering, University of Manchester, Stopford building, University of Manchester, 2002.

[76] R. H. Davies, T. F. Cootes, and C. J. Taylor. A minimum description length approach to statistical shape modelling. *Information Processing in Medical Imaging. 17th International Conference, IPMI 2001. Proceedings (Lecture Notes in Computer Science Vol.2082)*, pages 50–63, 2001.

[77] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. 3D statistical shape models using direct optimisation of description length. In *Proc. ECCV 2002*, volume 2352 of *LNCS*, pages 3–20. Springer, 2002. ISBN 3-540-43746-0.

[78] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. A minimum description length approach to statistical shape modeling. *Medical Imaging, IEEE Transactions on*, 21(5):525–537, 2002. ISSN 02780062.

[79] M. de Bruijne, B. van Ginneken, W. J. Niessen, and M. A. Viergever. Adapting active shape models for 3D segmentation of tubular structures in medical images. In *Information Processing in Medical Imaging*, volume 2732 of *Lecture Notes in Computer Science*, pages 136–147. Springer, 2003.

[80] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. John Wiley & Sons, 1998.

[81] A. Dubb, B. Avants, R. Gur, and J. Gee. Shape characterization of the corpus callosum in schizophrenia using template deformation. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, volume 2, pages 381–388, 2002.

[82] S. Duchesne, J. C. Pruessner, and D. L. Collins. Appearance-based modelling and segmentation of the hippocampus from MR images. *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, 3:2677–2680, 2001. ISSN 1094687x.

[83] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and Sons, New York, 2nd edition, 2001.

[84] J. S. Duncan and N. Ayache. Medical image analysis: progress over two decades and the challenges ahead. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1):85–106, 2000. ISSN 01628828.

[85] J. Duryea and J. M. Boone. A fully automatic algorithm for the segmentation of lung fields in digital chest radiographic images. *Medical Physics*, 22(2):183–191, 1995.

[86] J. P. Earls, V. B. Ho, T. K. Foo, E. Castillo, and S. D. Flamm. Cardiac MRI: Recent progress and continued challenges. *Journal of Magnetic Resonance Imaging*, 16(2):111 – 127, 2002.

[87] G. J. Edwards, T. F. Cootes, and C. J. Taylor. Advances in active appearance models. In *Proc. Int. Conf. on Computer Vision*, pages 137–142, 1999.

[88] G. J. Edwards, C. J. Taylor, and T. F. Cootes. Interpreting face images using active appearance models. In *Proc. 3rd IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 300–5. IEEE Comput. Soc, 1998.

[89] B. Efron. Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7:1–26, 1979.

[90] R. L. Ehman and J. P. Felmlee. Adaptive technique for high-definition MR imaging of moving structures. *Radiology*, 173(1):255–263, 1989.

[91] E. W. Ely and E. F. Haponik. Using the chest radiograph to determine intravascular volume status: the role of vascular pedicle width. *Chest*, 121(3):942–950, 2002.

[92] A. Ericsson. Automatic shape modelling and applications in medical imaging, Licentiate thesis, Centre for Mathematical Sciences, Lund University, 2004.

[93] SGI et al. Opengl. Web page: http://www.opengl.org/.

[94] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–95, 1981. ISSN 00010782.

[95] R. Fisker. *Making Deformable Template Models Operational*. PhD thesis, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, 2000.

[96] R. Fletcher. *Practical Methods of Optimization*. John Wiley & Sons, 1987.

[97] E. Forgey. Cluster analysis of multivariate data. *Biometrics*, 21:768, 1965.

[98] A. F. Frangi. *Three-dimensional model-based analysis of vascular and cardiac images*. PhD thesis, Image Sciences Institute, University Medical Center, Utrecht University, The Netherlands, 2001.

[99] A. F. Frangi, W. J. Niessen, and M. A. Viergever. Three-dimensional modeling for functional analysis of cardiac images: A review. *IEEE Transactions on Medical Imaging*, 20(1):2–25, 2001. ISSN 02780062.

[100] A. F. Frangi, D. Rueckert, and J. S. Duncan. Three-dimensional cardiovascular image analysis. *Medical Imaging, IEEE Transactions on*, 21(9):1005–1010, 2002. ISSN 02780062.

[101] A. F. Frangi, D. Rueckert, J. A. Schnabel, and W. J. Niessen. Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling. *IEEE Transactions on Medical Imaging*, 21(9):1151–1166, 2002. ISSN 02780062.

[102] C. M. Gallippi and E. T. Gregg. Automatic image registration for MR and ultrasound cardiac images. *Information Processing in Medical Imaging. 17th International Conference, IPMI 2001. Proceedings (Lecture Notes in Computer Science Vol.2082)*, pages 148–54, 2001.

[103] J. Gao, N. Ablitt, A. Elkington, and G. Z. Yang. Deformation modelling based on PLSR for cardiac magnetic resonance perfusion imaging. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2002. 5th International Conference. Proceedings, Part I (Lecture Notes in Computer Science Vol.2488)*, pages 612–19, 2002.

[104] M. Gerig, G. Jomier and M. Chakos. Valmet: a new validation tool for assessing and improving 3D object segmentation. In *MICCAI 2001*, number 2208 in Lecture Notes in Computer Science, pages 516–523. Springer, Berlin, 2001.

[105] N. A. Gershenfeld. *The Nature of Mathematical Modeling*. University Press, Cambridge, 1999.

[106] G. H. Golub and C. F. van Loan. *Matrix Computations*. Jonhs Hopkins, 1983.

[107] C. Goodall. Procrustes methods in the statistical analysis of shape. *Jour. Royal Statistical Society, Series B*, 53:285–339, 1991.

[108] J. C. Gower. Generalized Procrustes analysis. *Psychometrika*, 40:33–50, 1975.

[109] S. N. Gupta, M. Solaiyappan, G. M. Beache, A. E. Arai, and T. K. F. Foo. Fast method for correcting image misregistration due to organ motion in time-series MRI data. *Magnetic Resonance in Medicine*, 49(3):506–14, 2003. ISSN 07403194.

[110] D. L. Hall, G. S. Lodwick, R. P. Kruger, and S. J. Dwyer III. Computer diagnosis of heart disease. *Radiological Clinics of North America*, 9(3):533–541, 1971.

[111] D. W. Hansen, J. P. Hansen, M. Nielsen, A. S. Johansen, and M. B. Stegmann. Eye typing using Markov and active appearance models. In *IEEE Workshop on Applications of Computer Vision - WACV*, pages 132–136, dec 2002.

[112] D. W. Hansen, M. Nielsen, J. P. Hansen, A. S. Johansen, and M. B. Stegmann. Tracking eyes using shape and appearance. In *IAPR Workshop on Machine Vision Applications - MVA*, pages 201–204, dec 2002.

[113] T. Heap and D. Hogg. Extending the point distribution model using polar coordinates. In *Computer Analysis of Images and Patterns. 6th International Conference, CAIP'95*, pages 130–7. Springer-Verlag, 1995.

[114] T. Heap and D. Hogg. Improving specificity in PDMs using a hierarchical approach, 1997.

[115] K. B. Hilger, R. Larsen, and M. Wrobel. Growth modeling of human mandibles using non-Euclidean metrics. *Medical Image Analysis*, 7:425–433, 2003.

[116] K. B. Hilger and M. B. Stegmann. MADCam – the multispectral active decomposition camera. In *Proc. 10th Danish Conference on Pattern Recognition and Image Analysis, Copenhagen, Denmark*, volume 1, pages 136–142. DIKU, 2001.

[117] K. B. Hilger, M. B. Stegmann, and R. Larsen. A noise robust statistical model for image representation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2002, 5th Int. Conference, Tokyo, Japan*, volume 2488 of *LNCS*, pages 444–451, sep 2002.

[118] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46(3):R1–45, 2001. ISSN 00319155.

[119] P. W. Holland and R. E. Welsch. Robust regression using iteratively reweighted least-squares. *Commun. Statist.-Theor. Meth.*, A6:813–827, 1977.

[120] B. K. P. Horn, H. M. Hilden, and S. Negahdaripour. Closed-form solution of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America A (Optics and Image Science)*, 5 (7):1127–35, 1988. ISSN 07403232.

[121] J. L. Horn. A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30: 179–186, 1965.

[122] J. P. Hornak. The basics of MRI – an online textbook, 2004. URL http://www.cis.rit.edu/htbooks/mri/.

[123] X. Hou, S. Z. Li, H. J. Zhang, and Q. Cheng. Direct appearance models. *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 1:I–828, 2001.

[124] L. Hsu, K. L. Rhoads, A. H. Aletras, and A. E. Arai1. Surface coil intensity correction and non-linear intensity normalization improve pixel-resolution parametric maps of myocardial MRI perfusion. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003, 6th Int. Conference, Montréal, Canada*, LNCS 2879, pages 975–976. Springer, 2003.

[125] T. J. Hutton, B. R. Buxton, and P. Hammond. Dense surface point distribution models of the human face. *Proceedings IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2001)*, pages 153–60, 2001.

[126] A. Hyvarinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4-5):411–430, 2000. ISSN 08936080.

[127] Intel. Math kernel library. Web page: http://developer.intel.com/software/products/mkl/.

[128] J. Isidoro and S. Sclaroff. Active voodoo dolls: a vision based input device for nonrigid control. In *Proc. Computer Animation '98*, pages 137–143. IEEE Comput. Soc, 1998.

[129] C. Jackson, M. Robson, J. Francis, and J. A. Noble. Automatic planning of the acquisition of cardiac

mr images. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003, 6th Int. Conference, Montréal, Canada*, LNCS 2878, pages 541–548. Springer, 2003.

[130] A. K. Jain, Y. Zhong, and M.-P. Dubuisson-Jolly. Deformable template models: A review. *Signal Processing*, 71(2):109–129, 1998.

[131] A. K. Jain, Y. Zhong, and S. Lakshmanan. Object matching using deformable templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(3):267–278, 1996.

[132] P. Jannin, J. M. Fitzpatrick, D. J. Hawkes, X. Pennec, R. Shahidi, and M. W. Vannier. Validation of medical image processing in image-guided therapy. *IEEE Transactions on Medical Imaging*, 21(12): 1445–1449, 2002.

[133] M. J. Jones and T. Poggio. Multidimensional morphable models: a framework for representing and matching object classes. *International Journal of Computer Vision*, 29(2):107–31, 1998. ISSN 09205691.

[134] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. Jour. of Computer Vision*, 8(2):321–331, 1988.

[135] M. T. Kearney, J. Nolan, A. J. Lee, P. W. Brooksby, R. Prescott, A. M. Shah, A. G. Zaman, D. L. Eckberg, H. S. Lindsay, and P. D. Batin. A prognostic index to predict long-term mortality in patients with mild to moderate chronic heart failure stabilised on angiotensin converting enzyme inhibitors. *European Journal of Heart Failure*, 5(4):489–497, 2003.

[136] R. P. Kruger, J. R. Townes, D. L. Hall, S. J. Dwyer III, and G. S. Lodwick. Automated radiographic diagnosis via feature extraction and classification of cardiac size and shape descriptors. *IEEE Biomedical Transactions*, BME-19(3):174–186, 1972.

[137] R. Larsen. Shape modelling using maximum autocorrelation factors. In Ivar Austvoll, editor, *Proceedings of the Scandinavian Image Analysis Conference (SCIA'01)*, pages 98–103, Bergen, Norway, jun 2001.

[138] R. Larsen. Decomposition using maximum autocorrelation factors. *Journal of Chemometrics*, 16 (8-10):427–435, 2002.

[139] R. Larsen, H. Eiriksson, and M. B. Stegmann. Q-MAF shape decomposition. In Wiro J. Niessen and Max A. Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2001, 4th International Conference, Utrecht, The Netherlands*, volume 2208 of *LNCS*, pages 837–844. Springer, 2001.

[140] R. Larsen and K. B. Hilger. Statistical 2D and 3D shape analysis using non-Euclidean metrics. *Medical Image Analysis*, 7(4):417–423, 2003.

[141] R. Larsen, K. B. Hilger, K. Skoglund, S. Darkner, R. R. Paulsen, M. B. Stegmann, B. Lading, H. Thodberg, and H. Eiriksson. Some issues of biological shape modelling with applications. In J. Bigün and T. Gustavsson, editors, *13th Scandinavian Conference on Image Analysis (SCIA), Gothenburg, Sweden*, volume 2749 of *LNCS*, pages 509–519. Springer, jun 2003.

[142] R. Larsen, K. B. Hilger, and M. C. Wrobel. Statistical 2D and 3D shape analysis using non-Euclidean metrics. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2002, 5th Int. Conference, Tokyo, Japan*, Lecture Notes in Computer Science. Springer, 2002.

[143] Henrik B. W. Larsson, Thomas Fritz-Hansen, Egill Rostrup, Lars Søndergaard, Poul Ring, and Ole Henriksen. Myocardial perfusion modeling using MRI. *Magnetic Resonance in Medicine*, 35:716–726, 1996.

[144] B. Lelieveldt, R. van der Geest, S. C. Mitchell, J. Bosch, M. Sonka, and J. Reiber. 3D active appearance models: Fully automatic detection of endo- and epicardial contours in short-axis cardiac MR data. In *Proc. International Society of Magnetic Resonance In Medicine, ISMRM*, volume 2, page 1668. ISMRM, 2002.

[145] B. P. F. Lelieveldt. *Anatomical models in cardiovascular image analysis*. PhD thesis, Department of Radiology, Leiden University, The Netherlands, 1999.

[146] B. P. F. Lelieveldt, S. C. Mitchell, R. J. van der Geest, H. G. Bosch, M. Sonka, and J. H. C. Reiber.

Time continuous segmentation of cardiac MR images using active appearance motion models. In *Computer Assisted Radiology and Surgery, CARS 2001, International Congress Series*, volume 1230, pages 961–966. Elsevier Science, 2001.

[147] B. P. F. Lelieveldt, M. Uzumcu, R. J. van der Geest, J. H. C. Reiber, and M. Sonka. Multi-view active appearance models for consistent segmentation of multiple standard views: application to long- and short-axis cardiac MR images. *International Congress Series*, 1256:1141–1146, 2003. ISSN 05315131.

[148] B. P. F. Lelieveldt, R. J. van der Geest, H. J. Lamb, H. W. M. Kayser, and J. H. C. Reiber. Automatic, observer independent acquisition planning of cardiac MR short-axis views. In *Proc. International Society of Magnetic Resonance In Medicine – ISMRM 2000*, volume 1, page 200. ISMRM, 2000.

[149] D. C. Levin, L. Parker, J. H. Sunshine, and M. J. Pentecost. The practice of radiology – cardiovascular imaging: Who does it and how important is it to the practice of radiology? *American Journal of Roentgenology*, 178(2):303–306, 2002. ISSN 0361803x.

[150] L. Li, Y. Zheng, M. Kallergi, and R. A. Clark. Improved method for automatic identification of lung regions on chest radiographs. *Academic Radiology*, 8(7):629–638, 2001.

[151] M. Loog and B. van Ginneken. Supervised segmentation by iterated contextual pixel classification. In *Proceedings 16th International Conference on Pattern Recognition*, pages 925–928, 2002.

[152] A. Lundervold, N. Duta, T. Taxt, and A. K. Jain. Model-guided segmentation of corpus callosum in MR images. In *Computer Vision and Pattern Recognition*. IEEE Comput. Soc, 1999. ISBN 0769501494.

[153] H. Luo and T. O'Donnell. A 3D statistical shape model for the left ventricle of the heart. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2001, 4th International Conference, Utrecht, The Netherlands*, volume 2208 of *Lecture Notes in Computer Science*. Springer, 2001.

[154] T. Makela, P. Clarysse, O. Sipila, N. Pauna, Q. C. Pham, T. Katila, and I. E. Magnin. A review of cardiac image registration methods. *Medical Imaging, IEEE Transactions on*, 21(9):1011–1021, 2002. ISSN 02780062.

[155] I. Matthews and S. Baker. Active appearance models revisited. Technical Report CMU-RITR-03-02, Carnegie Mellon University Robotics Institute, 2003.

[156] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 2(1):91–108, 1996.

[157] M. F. McNitt-Gray, H. K. Huang, and J. W. Sayre. Feature selection in the pattern classification problem of digital chest radiograph segmentation. *IEEE Transactions on Medical Imaging*, 14(3): 537–547, 1995.

[158] Microsoft. Vision SDK, 2000. Web page: http://research.microsoft.com/projects/VisSDK/.

[159] S. Mitchell, B. Lelieveldt, R. Geest, H. Bosch, J. Reiber, and M. Sonka. Time continuous segmentation of cardiac MR image sequences using active appearance motion models. In *SPIE Medical Imaging 2001*, volume 1, pages 249–256, 2001.

[160] S. Mitchell, B. Lelieveldt, R. Geest, J. Schaap, J. Reiber, and M. Sonka. Segmentation of cardiac MR images: An active appearance model approach. In *Medical Imaging 2000: Image Processing, San Diego CA, SPIE*, volume 1, pages 224–234. SPIE, 2000.

[161] S. C. Mitchell, J. G. Bosch, B. P. F. Lelieveldt, R. J. van der Geest, J. H. C. Reiber, and M. Sonka. 3-D active appearance models: Segmentation of cardiac MR and ultrasound images. *Medical Imaging, IEEE Trans. on*, 21(9):1167–1178, 2002.

[162] S. C. Mitchell, B. P. Lelieveldt, J. G. Bosch, R. van der Geest, J. H. Reiber, and M. Sonka. Segmentation of cardiac MR volume data using 3D active appearance models. *SPIE Medical Imaging 2002*, pages 433–443, 2002.

[163] S. C. Mitchell, B. P. F. Lelieveldt, H. G. Bosch, J. H. C. Reiber, and M. Sonka. Disease characterization of active appearance model coefficients. *Proceedings of SPIE - The International Society for Optical Engineering*, 5032 II:949–957, 2003. ISSN 0277786X.

[164] S. C. Mitchell, B. P. F. Lelieveldt, R. J. van der Geest, H. G. Bosch, J. H. C. Reiver, and M. Sonka. Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images. *Medical Imaging, IEEE Transactions on*, 20(5):415–423, 2001. ISSN 02780062.

[165] L. Molgedey and H. G. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters*, 72(23):3634–3637, 1994.

[166] D. A. Moses and L. Axel. Quantification of 3-d curvature and shape of the interventricular septum. In *Proc. International Society of Magnetic Resonance In Medicine – ISMRM 2003*, volume 1, page 709. ISMRM, 2003.

[167] J. L. Mundy. Object recognition based on geometry: Progress over three decades. *Philosophical Transactions of the Royal Society London, Series A (Mathematical, Physical and Engineering Sciences)*, 356(1740):1213–1231, 1998.

[168] E. Nagel, N. Al-Saadiand, and E. Fleck. Cardiovascular magnetic resonance: Myocardial perfusion. *Herz*, 25(4):409–416, 2000. ISSN 03409937.

[169] E. Nagel, U. Schneider, S. Schalla, T. Ibrahim, B. Schnackenburg, A. Bornstedt, C. Klein, H. B. Lehmkuhl, and E. Fleck. Magnetic resonance real-time imaging for the evaluation of left ventricular function. *Journal of Cardiovascular Magnetic Resonance*, 2(1):7–14, 2000.

[170] E. Nagel, R. Underwood, D. Pennell, U. P. Sechtem, S. Neubauer, S. Wickline, O. M. Hess, M. Schwaiger, and E. Fleck. New developments in non-invasive cardiac imaging: critical assessment of the clinical role of cardiac magnetic resonance imaging. *European Heart Journal*, 19(9): 1286–1293, 1998.

[171] N. Nakamori, K. Doi, H. MacMahon, Y. Sasaki, and S. M. Montner. Effect of heart-size parameters computed from digital chest radiographs on detection of cardiomegaly: potential usefulness for computer-aided diagnosis. *Investigative Radiology*, 26(6):546–550, 1991.

[172] N. Nakamori, K. Doi, V. Sabeti, and H. MacMahon. Image feature analysis and computer-aided diagnosis in digital radiography: automated analysis of sizes of heart and lung in chest images. *Medical Physics*, 17(3):342–350, 1990.

[173] M. Nielsen, P. Johansen, A. D. Jackson, and B. Lautrup. Brownian warps: A least committed prior for non-rigid registration. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2002, 5th Int. Conference, Tokyo, Japan*, volume 2488 of *LNCS*, pages 557–564, sep 2002.

[174] C. R. Oost, B. P. F. Lelieveldt, M. Üzümcü, H. Lamb, J. H. C. Reiber, and M. Sonka. Multi-view active appearance models: Application to X-ray LV angiography and cardiac MRI. In *Information Processing in Medical Imaging*, volume 2732 of *LNCS*. Springer, 2003.

[175] E. Oost, B. P. F. Lelieveldt, G. Koning, M. Sonka, and J. H. C. Reiber. Left ventricle contour detection in X-ray angiograms using multi-view active appearance models. *SPIE Medical Imaging 2003*, 5032 I:394–404, 2003. ISSN 0277786X.

[176] X. Papademetris, A. J. Sinusas, D. P. Dione, R. T. Constable, and J.S. Duncan. Estimation of 3-d left ventricular deformation from medical images using biomechanical models. *Medical Imaging, IEEE Transactions on*, 21(7):786–800, 2002. ISSN 02780062.

[177] A. Pednekar, I. A. Kakadiaris, R. Muthupillai, and S. Flamm. Validation of the automatic computation of the ejection fraction from cine-MRI. In *MICCAI, 6th Int. Conference, Montréal, Canada*, LNCS 2879, pages 987–989. Springer, 2003.

[178] D. Perperidis, A. Rao, M. Lorenzo-Valdés, R. Mohiaddin, and D. Rueckert. Spatio-temporal alignment of 4D cardiac MR images. In *Functional Imaging and Modeling of the Heart, FIMH 2003*, volume 2674 of *LNCS*, pages 205–214, Lyon, France, 2003. Springer.

[179] E. Pietka. Lung segmentation in digital chest radiographs. *Journal of Digital Imaging*, 2:79–84, 1994.

[180] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. R. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, second edition, 1992.

[181] J. Price, Y. Rogers, H. Sharp, D. Benyon, S. Holland, and T. Carey, editors. *Human-Computer Inter-*

action. Addison-Wesley, 1994.

[182] A. Rao, G. I. Sanchez-Ortiz, R. Chandrashekara, M. Lorenzo-Valdés, R. Mohiaddin, and D. Rueckert. Construction of a cardiac motion atlas from MR using non-rigid registration. In *Functional Imaging and Modeling of the Heart, FIMH 2003*, volume 2674 of *LNCS*, pages 205–214, Lyon, France, 2003. Springer.

[183] J. A. Richards and X. Jia. *Remote sensing digital image analysis: an introduction*. Springer Verlag, 3rd edition, 1999.

[184] M. G. Roberts, T. F. Cootes, and J. E. Adams. Linking sequences of active appearance sub-models via constraints: an application in automated vertebral morphometry. In *Proceedings of the British Machine Vision Conference, BMVC*, volume 1, pages 349–358, 2003.

[185] M. Rogers. *Exploiting Weak Constraints on Object Structure and Appearance for Segmentation of 2-D Images*. PhD thesis, University of Manchester, 2001.

[186] S. Romdhani, S. Gong, and A. Psarrou. A multi-view nonlinear active shape model using kernel PCA. *BMVC99. Proceedings of the 10th British Machine Vision Conference*, pages 483–92 vol.2, 1999.

[187] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *Medical Imaging, IEEE Transactions on*, 18(8):712–721, 1999. ISSN 02780062.

[188] S. Sclaroff and J. Isidoro. Active blobs. *Proc. of the Int. Conf. on Comput. Vision*, pages 1146–1153, 1998.

[189] S. Sclaroff and A. P. Pentland. Modal matching for correspondence and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):545–61, 1995.

[190] I. M. Scott, T. F. Cootes, and C. J. Taylor. Improving appearance model matching using local image structure. In *Proc. Information Processing in Medical Imaging, IPMI 2003*, volume 2878 of *LNCS*, pages 258–269, 2003.

[191] M. Sermesant, C. Forest, X. Pennec, H. Delingette, and N. Ayache. Deformable biomechanical models: Application to 4d cardiac image analysis. *Medical Image Analysis*, 7(4):475–488, 2003. ISSN 13618415.

[192] J. A. Sethian. *Level Set Methods and Fast Marching*. Cambrige University Press, 1999. 2nd edition.

[193] N. Sezaki and K. Ukena. Automatic computation of the cardiothoracic ratio with application to mass screening. *IEEE Transactions on Biomedical Engineering*, BME-20(4):248–253, 1973.

[194] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *American Journal of Roentgenology*, 174:71–74, 2000.

[195] H. Si. Tetgen – A 3D Delaunay tetrahedral mesh generator. Technical Report 4, Weierstrass Institute for Applied Analysis and Stochastics, WIAS, Berlin, 2002. URL http://tetgen.berlios.de.

[196] N. P. Smith, P. J. Mulquiney, M. P. Nash, C. P. Bradley, D. P. Nickerson, and P. J. Hunter. Mathematical modelling of the heart: cell to organ. *Chaos, Solitons and Fractals*, 13(8):1613–1621, 2002. ISSN 09600779.

[197] M. Sonka, J. G. Bosch, B. P. F. Lelieveldt, S. C. Mitchell, and J. H. C. Reiber. Computer-aided diagnosis via model-based shape analysis: cardiac MR and echo. *International Congress Series*, 1256:1013–1018, 2003. ISSN 05315131.

[198] M. Sonka, B. P. F. Lelieveldt, S. C. Mitchell, J. G. Bosch, R. J. van der Geest, and J. H. C. Reiber. Active appearance motion model segmentation. *Proceedings Second International Workshop on Digital and Computational Video*, pages 64–68, 2001.

[199] P. D. Sozou, T. F. Cootes, C. J. Taylor, and E. C. Di Mauro. Non-linear generalization of point distribution models using polynomial regression. *Image and Vision Computing*, 13(5):451–7, 1995.

[200] L. Spreeuwers and M. Breeuwer. Automatic detection of the myocardial boundaries of the right and left ventricle in MR cardio perfusion scans. *Proceedings of SPIE - The International Society for Optical Engineering*, 4322(3):1207–1217, 2001. ISSN 0277786X.

[201] L. J. Spreeuwers, F. Wierda, and M. Breeuwer. Optimal myocardial boundary estimation for MR cardio perfusion measurements using sensitivity analysis. *Computers in Cardiology, 2002*, pages 197–200, 2002. ISSN 02766547.

[202] L. H. Staib and J. S. Duncan. Boundary finding with parametrically deformable models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14(11):1061–1075, 1992. ISSN 01628828.

[203] M. B. Stegmann. Active appearance models: Theory, extensions and cases. Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, 2000. http://www.imm.dtu.dk/∼aam/.

[204] M. B. Stegmann. Analysis of 4D cardiac magnetic resonance images. *Journal of The Danish Optical Society, DOPS-NYT*, 4:38–39, dec 2001.

[205] M. B. Stegmann. Object tracking using active appearance models. In *Proc. 10th Danish Conference on Pattern Recognition and Image Analysis, Copenhagen, Denmark*, volume 1, pages 54–60. DIKU, 2001.

[206] M. B. Stegmann. Analysis and segmentation of face images using point annotations and linear subspace techniques. Technical Report IMM-REP-2002-22, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, aug 2002.

[207] M. B. Stegmann. An annotated dataset of 14 cardiac MR images. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, apr 2002.

[208] M. B. Stegmann. An annotated dataset of 14 meat images. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, apr 2002.

[209] M. B. Stegmann. The AAM-API: An open source active appearance model implementation. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003, 6th Int. Conference, Montréal, Canada*, LNCS 2879, pages 951–952. Springer, nov 2003.

[210] M. B. Stegmann. Bi-temporal 3D active appearance modelling with applications to unsupervised ejection fraction estimation from 4D cardiac MRI. 2004 (to be submitted).

[211] M. B. Stegmann and R. H. Davies. Automated analysis of corpora callosa. Technical Report IMM-REP-2003-02, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, mar 2003.

[212] M. B. Stegmann, R. H. Davies, and C. Ryberg. Corpus callosum analysis using MDL-based sequential models of shape and appearance. In *International Symposium on Medical Imaging 2004, San Diego CA, SPIE*. SPIE, feb 2004 (in press).

[213] M. B. Stegmann, B. K. Ersbøll, and R. Larsen. FAME – a flexible appearance modelling environment. *IEEE Trans. on Medical Imaging*, 22(10):1319–1331, 2003.

[214] M. B. Stegmann, R. Fisker, and B. K. Ersbøll. Extending and applying active appearance models for automated, high precision segmentation in different image modalities. In *Proc. 12th Scandinavian Conference on Image Analysis – SCIA 2001*, volume 1, pages 90–97, 2001.

[215] M. B. Stegmann, R. Fisker, B. K. Ersbøll, H. H. Thodberg, and L. Hyldstrup. Active appearance models: Theory and cases. In *Proc. 9th Danish Conference on Pattern Recognition and Image Analysis, Aalborg, Denmark*, volume 1, pages 49–57. AUC, 2000.

[216] M. B. Stegmann and S. Forchhammer. On decomposing object appearance using PCA and wavelet bases with applications to image segmentation. In Hans Joachim Werner, editor, *MATRIX'02, Eleventh International Workshop on Matrices and Statistics*, page 21. Informatics and Mathematical Modelling, Technical University of Denmark, DTU, sep 2002.

[217] M. B. Stegmann and S. Forchhammer. On exploiting wavelet bases in statistical region-based segmentation. In *Proc. 11th Danish Conference on Pattern Recognition and Image Analysis*, volume 1, pages 75–82, Copenhagen, Denmark, aug 2002. DIKU.

[218] M. B. Stegmann, S. Forchhammer, and T. F. Cootes. Wavelet enhanced appearance modelling. In *International Symposium on Medical Imaging 2004, San Diego CA, SPIE*. SPIE, 2004 (in press).

[219] M. B. Stegmann and D. D. Gomez. A brief introduction to statistical shape analysis. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, mar 2002.

[220] M. B. Stegmann and R. Larsen. Multi-band modelling of appearance. In *First International Workshop on Generative Model-Based Vision – GMBV, ECCV 2002*, pages 101–106, Copenhagen, Denmark, jun 2002. DIKU.

[221] M. B. Stegmann and R. Larsen. Multi-band modelling of appearance. *Image and Vision Computing*, 21(1):61–67, jan 2003.

[222] M. B. Stegmann, R. Larsen, and H. B. W. Larsson. Unsupervised correction of physiologically-induced slice-offsets in 4D cardiac MRI. *Journal of Cardiovascular Magnetic Resonance (7th Annual SCMR Meeting, Barcelona, Spain)*, 6(1):451–452, 2004.

[223] M. B. Stegmann and H. B. W. Larsson. Fast registration of cardiac perfusion MRI. In *Proc. International Society of Magnetic Resonance In Medicine – ISMRM 2003, Toronto, Ontario, Canada*, page 702, Berkeley, CA, USA, 2003. ISMRM.

[224] M. B. Stegmann and H. B. W. Larsson. Motion-compensation of cardiac perfusion MRI using a statistical texture ensemble. In *Functional Imaging and Modeling of the Heart, FIMH 2003*, volume 2674 of *LNCS*, pages 151–161, Lyon, France, 2003. Springer.

[225] M. B. Stegmann and H. B. W. Larsson. Rapid and unsupervised correction of respiratory-induced motion in 4D cardiac cine MRI. 2004 (to be submitted).

[226] M. B. Stegmann, J. C. Nilsson, and B. A. Grønning. Automated segmentation of cardiac magnetic resonance images. In *Proc. International Society of Magnetic Resonance In Medicine – ISMRM 2001*, volume 2, page 827. ISMRM, 2001.

[227] M. B. Stegmann, H. Ólafsdóttir, and H. B. W. Larsson. Unsupervised motion-compensation of multi-slice cardiac perfusion MRI. *Invited contribution for the FIMH special issue in Medical Image Analysis*, 2004 (submitted).

[228] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999.

[229] P. Switzer. Min/max autocorrelation factors for multivariate spatial imagery. In L. Billard, editor, *Computer Science and Statistics*, pages 13–16. Elsevier Science Publishers B.V., 1985.

[230] G. Szekely, A. Kelemen, C. Brechbuhler, and G. Gerig. Segmentation of 2-d and 3-d objects from mri volume data using constrained elastic deformations of flexible fourier contour and surface models. *Medical Image Analysis*, 1(1):19–34, 1996. ISSN 13618415.

[231] J. M. F. ten Berge. Orthogonal procrustes rotation for two or more matrices. *Psychometrika*, 42: 267–276, 1977.

[232] N. A. Thacker, A. Jackson, D. Moriarty, and E. Vokurka. Improved quality of re-sliced MR images using re-normalized sinc interpolation. *Journal of Magnetic Resonance Imaging*, 10(4):582–588, 1999. ISSN 10531807.

[233] The Economist. Medical imaging – MRI's inside story, Dec. 4th 2003.

[234] H. H. Thodberg. Hands-on experience with active appearance models. In *SPIE Medical Imaging 2002*, pages 495–506. SPIE, 2002.

[235] H. H. Thodberg. Minimum description length shape and appearance models. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2003.

[236] O. Tsujii, M.T. Freedman, and S.K Mun. Automated segmentation of anatomic regions in chest radiographs using an adaptive-sized hybrid neural network. *Medical Physics*, 25(6):998–1007, 1998.

[237] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proc. 1991 IEEE Com. Soc. Conf. on CVPR*, pages 586–91. IEEE Com. Soc. Press, 1991.

[238] C. J. Twining and S. Marsland. Constructing diffeomorphic representations of non-rigid registrations of medical images. In *Information Processing in Medical Imaging, IPMI, Ambleside, UK*, IPMI, 2003.

[239] C. J. Twining and C. J. Taylor. Kernel principal component analysis and the construction of non-linear active shape models. In *Proceedings of the British Machine Vision Conference, BMVC*, volume 1, pages 23–32, 2001.

[240] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 13(4):376–380, 1991. ISSN 01628828.

[241] M. Üzümcü, A. F. Frangi, J. H. C. Reiber, and B. P. F. Lelieveldt. Independent component analysis in statistical shape models. In *Proceedings of SPIE - The International Society for Optical Engineering*, volume 5032 I, pages 375–383, 2003.

[242] M. Üzümcü, A. F. Frangi, M. Sonka, J. H. C. Reiber, and B. P. F. Lelieveldt. ICA vs. PCA active appearance models: Application to cardiac MR segmentation. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003, 6th Int. Conference, Montréal, Canada*, volume 2878 of *LNCS*, pages 451 – 458, 2003.

[243] B. van Ginneken, A. F. Frangi, J. J. Staal, B. M. ter Haar Romeny, and M. A. Viergever. Active shape model segmentation with optimal features. *IEEE Transactions on Medical Imaging*, 21(8):924–933, 2002.

[244] B. van Ginneken, S. Katsuragawa, B. M. ter Haar Romeny, K. Doi, and M. A. Viergever. Automatic detection of abnormalities in chest radiographs using local texture analysis. *IEEE Transactions on Medical Imaging*, 21(2):139–149, 2002.

[245] B. van Ginneken, M. B. Stegmann, and M. Loog. Segmentation of anatomical structures in chest radiographs using supervised methods: A comparative study on a public database. *Medical Image Analysis*, 2004 (submitted).

[246] B. van Ginneken and B. M. ter Haar Romeny. Automatic segmentation of lung fields in chest radiographs. *Medical Physics*, 27(10):2445–2455, 2000.

[247] N. F. Vittitoe, R. Vargas-Voracek, and C. E. Floyd Jr. Identification of lung regions in chest radiographs using Markov Random Field modeling. *Medical Physics*, 25(6):976–985, 1998.

[248] C.-F. Westin, J. Richolt, V. Moharir, and R. Kikinis. Affine adaptive filtering of CT data. *Medical Image Analysis*, 4(2):161–177, 2000. ISSN 13618415.

[249] C. B. H. Wolstenholme and C. J. Taylor. Using wavelets for compression and multiresolution search with active appearance models. *BMVC99. Proceedings of the 10th British Machine Vision Conference*, pages 473–82 vol.2, 1999.

[250] C. B. H. Wolstenholme and C. J. Taylor. Wavelet compression of active appearance models. In *Medical Image Computing and Computer-Assisted Intervention, MICCAI*, pages 544–554, 1999.

[251] C. B. H. Wolstenholme and C. J. Taylor. Wavelet compression of active appearance models. In *Medical Image Understanding and Analysis, Oxford*, page 4, 1999.

[252] X. W. Xu and K. Doi. Image feature analysis for computer-aided diagnosis: accurate determination of ribcage boundary in chest radiographs. *Medical Physics*, 22(5):617–626, 1995.

[253] X. W. Xu and K. Doi. Image feature analysis for computer-aided diagnosis: detection of right and left hemidiaphragm edges and delineation of lung field in chest radiographs. *Medical Physics*, 23 (9):1613–1624, 1996.

[254] G. Z. Yang, P. D. Gatehouse, J. Panting, P. Burger, D. J. Pennell, and D. N. Firmin. Motion analysis for magnetic resonance myocardial perfusion imaging. *Image Processing and Its Applications, 1997., Sixth International Conference on*, 2:838–842 vol.2, 1997.

[255] A. L. Yuille, P. W. Hallinan, and D. S. Cohen. Feature extraction from faces using deformable templates. *Int. Jour. of Computer Vision*, 8(2):99–111, 1992.