



TECHNISCHE  
UNIVERSITÄT  
WIEN  
  
VIENNA  
UNIVERSITY OF  
TECHNOLOGY

DIPLOMARBEIT

# Automatic Assessment of the Knee Alignment Angles on Full-limb Radiographs

ausgeführt am Institut für  
Computergestützte Automation,  
Abteilung für Mustererkennung und Bildverarbeitung  
der Technischen Universität Wien

unter der Anleitung von  
Ao. Univ. Prof. Dipl.-Ing. Dr.techn. Robert Sablatnig  
und  
Dr.techn. Georg Langs

durch

**Bakk.techn. Peter Andreas Widhalm**

Matrikelnummer: 9926313

Toßgasse 4/28

1150 Wien.

Wien, 21. Oktober 2008

---

## Abstract

The knee alignment angle, defined by the mechanical axes of the femur and the tibia, is of high importance in orthopedics and traumatology, in particular for pre-operative planning and post-operative follow-up assessment. It serves as a predictor for pre-arthritis and the post-operative angulation is known to be associated with the clinical outcome.

So far, only manual and semi-automated methods of measuring the alignment angle exist. The most serious drawback of these techniques is that they lack reproducibility. The points in the anatomical structure, which define the angle to be measured, are only vaguely defined and cannot clearly and precisely be identified. The resulting variability between repeated measurements precludes the detection of small changes. A fully automatic measurement method of axis alignment that provides a consistent definition of anatomical landmarks would eliminate inter and intra reader variabilities caused by human interpretation and lead to more accurate and reproducible results.

In the course of this thesis, a novel method for the automatic measurement of alignment angles is developed and prospectively tested. It allows a fully automatic assessment of knee alignment angles in full-limb radiographs with high precision.

The positions of the lower limb bones and joints are estimated using *Sparse Markov Random Field Appearance Models*. They are able to detect anatomical structures by configurations of interest points, taking their spatial arrangement and local appearance into account.

Based on the coarse position estimates of the bones, their contours are delineated by *Active Shape Models*, controlled by ongoing estimates of the reliability of the model. The regions around the joints are refined using submodels. Landmarks are identified by their index and can be matched between different instances of a shape. Hence, the defining points of the axes can be located in a straightforward and repeatable way, when they are directly represented by landmarks on the contour and annotated manually in the training phase on any instance of the bone.

Overlapping structures and the compound nature of the large radiography acquisition which results in partially missing data and changing intensities let standard ASMs fail. For this reason, a search procedure is introduced, which is controlled by ongoing estimates of the fit confidence during the search, leading to an improved result robustness, even if the spatial initialization is poor and the structures of interest are partially cropped or occluded.

Experimental results show that the automatic assessment of the knee alignment angle allows for an accurate and observer independent quantification with high precision and improves the detection of small changes.

## Zusammenfassung

Der Beinachsen-Winkel, der durch die mechanischen Achsen des Femur und der Tibia definiert wird, ist von großer Bedeutung für die Orthopädie und Traumatologie, insbesondere für die Operationsplanung und die post-operative Nachversorgung. Ferner dient er als Prädiktor von Präarthritiden. Post-operative Messungen geben Auskunft über den Erfolg des Eingriffs.

Bisher gibt es ausschließlich manuelle und semi-automatisierte Meßmethoden. Ihr größter Nachteil liegt in der mangelhaften Reproduzierbarkeit der Ergebnisse. Der zu messende Winkel wird durch Punkte in der anatomischen Struktur bestimmt, die nur vage definiert sind und nicht präzise identifiziert werden können. Die resultierende Variabilität zwischen wiederholten Messungen macht die Erkennung von kleinen Änderungen unmöglich. Eine voll-automatische Methode, die auf klar definierten Messpunkten basiert, würde die durch menschliche Interpretation verursachte Variabilität zwischen wiederholten Messungen beseitigen und zu akkurateren und reproduzierbareren Ergebnissen führen.

Im Zuge dieser Arbeit wurde eine neue Methode entwickelt und prospektiv getestet, die eine voll-automatische Messung des Beinachsen-Winkels auf Ganzbein-Röntgenaufnahmen mit hoher Präzision erlauben soll. Die Positionen der Beinknochen und -gelenke werden grob durch *Sparse Markov Random Field Appearance Models* ermittelt. Diese können anatomische Strukturen durch Konfigurationen von "interest points" erkennen, indem sie ihre räumliche Anordnung und lokale Erscheinung betrachten.

Basierend auf der groben Lokalisierung der Knochen werden deren Konturen durch *Active Shape Models* erkannt. Dabei wird die Suche von laufenden Schätzungen der Zuverlässigkeit der aktuellen Modellanpassung gesteuert. Die Regionen um die Gelenke werden durch die Verwendung von Submodellen verfeinert. Die Modell-Punkte können durch ihren Index identifiziert und zwischen unterschiedlichen Instanzen einer Form zugeordnet werden. Daher können die Punkte, die zur Bestimmung der Achsen dienen, in einfacher Weise ermittelt werden, indem sie direkt durch Modell-Punkte auf der Kontur repräsentiert und während der Trainings-Phase manuell annotiert werden.

Durch die aus mehreren Teilen zusammengesetzten großen Röntgen-Aufnahmen und durch überlappende Strukturen ist die Bildinformation teilweise unvollständig und mit variierender Intensität behaftet, wodurch die übliche ASM Segmentierung oftmals fehlschlägt. Aus diesem Grund wird eine Segmentierungsmethode vorgestellt, die durch laufende Schätzungen der "Fit Confidence" (Anpassungs-Qualität) gesteuert wird und zu verbesserter Robustheit der Ergebnisse führt, auch wenn die räumliche Initialisierung ungenau und die für die Messung relevanten Strukturen teilweise abgeschnitten oder verdeckt sind.

Experimentelle Ergebnisse zeigen, dass die automatische Messung des Beinachsen-Winkels eine präzise und beobachterunabhängige Quantifizierung ermöglicht und die Erkennung kleiner Änderungen verbessert.

# Synopsis

Chapter 1 gives an introduction which highlights the clinical background of this work, outlines the state of the art in assessing knee alignment and states the aim of this thesis. The remaining chapters divide the thesis into three parts.

In chapter 2 the methods deployed for image interpretation are explained. It starts with an overview of the prevalent techniques and then focuses on a sequence of concrete methods capable of delineating the lower limb bones and measuring the alignment angles in a fully automatic way. The description of the methods for the acquisition of training samples is followed by an illustration of how shape and appearance can be modeled statistically. *Sparse Markov Random Field Appearance Models* and, as an alternative, an application specific approach are presented as techniques for coarsely localizing the bones on a target image. The resulting estimations are used to initialize an *Active Shape Model* (ASM) search, which is explained in detail. The remainder of this chapter points out the shortcomings of the standard ASM search and introduces a measure of *fit confidence*, which can be utilized to both express the confidence in the results and improve accuracy and result robustness, by controlling the search procedure by ongoing estimates of the fit confidence.

Chapter 3 reports on the experimental evaluation of the new methods. The power and specificity of failure detection based on *fit confidence* is assessed and the accuracy of the results are compared between the standard ASM search and the enhanced search procedure. Further, the automatic assessment of the knee alignment angle is evaluated functionally. The agreement of the results with the standard of reference is calculated to determine accuracy. To evaluate repeatability of the results, the agreement between repeated measurements is computed. The findings are benchmarked against the results of manual measuring techniques.

Parts of the work performed during this thesis also contributed to the *Workshop of the Austrian Association for Pattern Recognition, OAGM/AAPR 2008* [WLD<sup>+</sup>08]. The paper outlines the methods introduced to increase result robustness, described in Chapter 2, and summarizes the experimental results presented in Chapter 3.

# Acknowledgments

I would like to express my thankfulness to everyone who helped shaping this thesis by sharing their expertise and experience with me, by providing me insights and knowledge about technical and medical questions, by discussing arising issues and challenges of theoretical and practical nature with me, by proofreading, by being patient and understanding or by supporting me in any other way.

I highly appreciated working with Negar Fakhrai and Philipp Peloschek from the Department of Diagnostic Radiology of the Vienna General Hospital (AKH Wien). I thank Georg Langs and Rene Donner for providing me with literature and useful MATLAB source code, for giving me precious hints and allowing me to discuss my topic with them and also for their help with organizational matters. Further, I would like to thank Professor Robert Sablatnig for supervision.

Most of all, I would like to thank and to apologize to Mimi, who had to suffer through all this with me.

The image data needed for the studies in this thesis was provided by the Department of Diagnostic Radiology at the AKH Wien.

# Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	The role of knee alignment . . . . .	7
1.2	Definition of the knee alignment angle . . . . .	7
1.3	Conventional measuring techniques . . . . .	8
1.4	The aim of this thesis . . . . .	10
<b>2</b>	<b>Automatic image interpretation</b>	<b>11</b>
2.1	Obtaining the standard of reference data . . . . .	12
2.1.1	Landmarks . . . . .	13
2.1.2	Delineating the bones with Live-wire . . . . .	14
2.1.3	The training instances . . . . .	16
2.2	Active Shape Models . . . . .	16
2.2.1	Modelling shape . . . . .	17
2.2.1.1	Shape and Pose . . . . .	17
2.2.1.2	Alignment of the training set . . . . .	18
2.2.1.3	Principal Component Analysis . . . . .	19
2.2.1.4	Statistical Model of Shape Variation . . . . .	21
2.2.1.5	Automatically establishing landmark correspondences with the Minimum Description Length Approach . . . . .	24
	Parameterization . . . . .	25
2.2.2	Modelling local texture . . . . .	28
2.2.3	Multi-resolution Active Shape Models . . . . .	30
2.3	Coarse localization of the bones . . . . .	30
2.3.1	An application-specific approach . . . . .	31
2.3.2	Sparse Markov Random Field Appearance Models . . . . .	32
2.4	Image interpretation with standard Active Shape Models . . . . .	34
2.4.1	Updating the landmark positions according to image content	35
2.4.2	Updating pose and shape parameters . . . . .	36

2.5	Improvements to the ASM Approach . . . . .	36
2.5.1	Related work . . . . .	37
2.5.2	Estimating Fit Confidence . . . . .	38
	Search procedure . . . . .	38
2.5.3	Using Fit Confidence to control search . . . . .	41
2.5.3.1	Termination of the search procedure . . . . .	42
2.5.3.2	Variable search range . . . . .	44
2.5.3.3	Variable deformation limits and weights for shape alignment . . . . .	45
2.5.3.4	Re-initialization by interposition of an alterna- tive search strategy . . . . .	47
2.5.4	Using sub-models and super-models . . . . .	47
2.6	Identifying the measuring points and calculating the axis angle .	50
2.7	Summary of the image interpretation process . . . . .	50
<b>3</b>	<b>Experiments</b>	<b>54</b>
3.1	Common Setup . . . . .	54
3.2	Fit Confidence . . . . .	54
3.2.1	Experimental setup . . . . .	54
3.2.2	Results . . . . .	55
3.2.3	Discussion . . . . .	57
3.3	Search performance . . . . .	57
3.3.1	Experimental setup . . . . .	57
3.3.2	Results . . . . .	58
3.3.3	Discussion . . . . .	59
3.4	Angle measurement . . . . .	59
3.4.1	Experimental setup . . . . .	59
3.4.2	Results . . . . .	60
3.4.3	Discussion . . . . .	62
<b>4</b>	<b>Conclusion</b>	<b>69</b>

# List of Figures

1.1	Full leg radiograph and definition of the knee alignment angle . . .	9
2.1	The contours of the right femoral bone captured with Live-Wire	15
2.2	The principal axes of a two-dimensional data set. . . . .	20
2.3	Covariance matrix of 13 right femoral bones with 256 landmarks each. . . . .	21
2.4	Correlation matrix of 13 right femoral bones with 256 landmarks each. . . . .	22
2.5	Deformation of the mean shape using the first three modes . . .	23
2.6	The decrease of eigenvalues of the training set of right femoral bones . . . . .	24
2.7	The first three modes of variation of a model built from the training set of outlines of the left femoral bone with arc-length parametrization and MDL parametrization . . . . .	28
2.8	Gray value profiles orthogonal to the shape. . . . .	29
2.9	Gray value samples $g_j^i$ along the profile. . . . .	29
2.10	Normalized derivative of gray value samples $g_j^i$ along the profile.	29
2.11	An application specific method to coarsely localize the bones . .	31
2.12	Examples of Gradient Vector Flow with the detected symmetry interest points . . . . .	33
2.13	Descriptor extraction from the GVF field . . . . .	33
2.14	MRF graph . . . . .	34
2.15	The shape error function $e_S$ . . . . .	40
2.16	Fit Confidence calculated locally for individual landmarks . . . .	42
2.17	Euclidean distance between the landmark coordinate vectors of two successive iterations of the search procedure. . . . .	43
2.18	The maximum euclidean distance of a landmark to its correct position over the iterations of the search procedure. . . . .	43
2.19	1 - Fit Confidence over the iterations of the search procedure. . .	43
2.20	Challenges in the automatic interpretation of the full-limb radio- graphs . . . . .	45



2.21	Shortcomings of the local texture matching method . . . . .	46
2.22	The cost map used for re-initialization . . . . .	48
2.23	Translation coefficients resulting from the cheapest path through the cost map . . . . .	49
2.24	Prototypical instance of the right femor, generated by setting the model parameters to their mean values. . . . .	51
2.25	Prototypical instance of the right femoral head with the three points needed to calculate its center. . . . .	52
2.26	The center of the intercondular notch annotated on a prototypical instance of the proximal side of the right knee joint. . . . .	52
2.27	The center of the tibial spines annotated on a prototypical in- stance of the distal side of the right knee joint. . . . .	52
2.28	The center of the ankle mortise annotated on a prototypical in- stance of the right ankle joint. . . . .	52
2.29	Schematic illustration of the image interpretation process . . . .	53
3.1	The distribution of the mean euclidean distance of the landmarks to their true positions given $F > t_F$ with varied thresholds $t_F$ of Fit Confidence. . . . .	55
3.2	The distribution of the mean euclidean distance of the landmarks to their true positions given $F \leq t_F$ with varied thresholds $t_F$ of Fit Confidence. . . . .	56
3.3	The error rates for a distance threshold $t_d = 10$ and varying thresholds $t_F$ of Fit Confidence . . . . .	56
3.4	The mean landmark errors resulting from the different algorithm variants. . . . .	58
3.5	Distribution of the landmark errors resulting from the different algorithm variants. . . . .	58
3.6	Results of the validation set: automatically assessed results plot- ted against standard of reference, with line of equality . . . . .	61
3.7	Results of the validation set: difference between automatic mea- surement and standard of reference, plotted against their mean, with bias and limits of agreement . . . . .	61
3.8	Results of the validation set: difference between automatic mea- surements with varied clipping of the radiograph against their mean . . . . .	62
3.9	Results of the LOOCV: automatically assessed results plotted against standard of reference, with line of equality . . . . .	67
3.10	Results of the LOOCV: difference between automatic measure- ment and standard of reference, plotted against their mean, with bias and limits of agreement . . . . .	67
3.11	Results of the LOOCV: difference between automatic measure- ments with varied clipping of the radiograph against their mean .	68

# Chapter 1

## Introduction

### 1.1 The role of knee alignment

The assessment of lower limb alignment is important in the field of orthopedics and traumatology, particularly when planning surgery such as total knee arthroplasty [LE77, PT92, PE88]. The knee alignment is of utmost importance in the planning of corrective knee surgery and the degree of postoperative angulation is known to be directly associated with the clinical outcome [CIW93, HD04]. In addition, the degree of mal-alignment correlates with the degree of cartilage loss assessed by magnetic resonance tomography [CWHW04]. It is known that quantitative joint space loss detected on radiography correlates with possible prognostic value in knee osteoarthritis (OA) [SSF<sup>+</sup>01] and knee alignment angle is closely related with structural progression in knee OA and with functional decline [CWHW04, GB07]. OA is a wide spread pathology which is commonly associated with deformities of the alignment of the lower limb [CSL<sup>+</sup>97]. In this context, the varus knee alignment measurement focuses on arthritic changes in the medial knee compartment whereas the valgus knee alignment affects the lateral knee compartment. Further, the alignment of the lower extremity can be thought of as a predictor for pre-arthritis and also plays a significant role in the postoperative follow-up.

### 1.2 Definition of the knee alignment angle

Many authors suggest using full-limb radiographs to measure the hip-knee-ankle angle (mechanical axis angle) to plan high tibial osteotomy, e.g. [BA00], [Gre95], [Mur94]. In [IDC<sup>+</sup>07] the anatomic-axis angle (the femur-tibia angle) measured from knee radiographs is suggested to assess alignment, in order to avoid pelvic radiation and the need for special equipment. [CLS94], [YCTN07] and others define the mechanical axis angle as the angle between the mechanical axis of the femur and that of the tibia, i.e. the angle between two lines, each defined by two points. Others, e.g. [BA00] and [HFB<sup>+</sup>06], define this angle by only three points: the center of the femoral head, the center of the knee and the center of the ankle mortise. The center of the femoral head was defined as the

center of the circumscribed circle of a triangle formed by three points along the medial part of the femoral head, dividing it into portions of roughly equal size. The center of the knee is defined to be in the middle between the intercondylar notch and the center of the tibial spines. Considering a limited accuracy for automatically localizing these two points, the latter approach is expected to yield more stable angle measurements than trying to measure the angle between two lines, each meeting one of these points exactly. The angles resulting from the two approaches - assuming exact localization - deviate acceptably little. Therefore, the three-point approach was adopted in this work, as depicted in Fig. 1.1.

### 1.3 Conventional measuring techniques

Conventional measuring methods are the physical examination of the anatomic-axis angle using a goniometer and radiographic assessment of the mechanical axis, either on conventional or digital radiographs. The angle formed by the intersection of the lines drawn from the hip to the knee and from the knee to the ankle on a radiograph of the entire lower extremity is currently considered the gold standard measure of knee alignment [KVWM05]. A standardized protocol stipulates standing antero-posterior long-leg radiographs, using a conventional film screen system [SSP<sup>+</sup>04, MBH87].

However, full-limb radiographs are limited in that they lack standardisation of the precise positioning of the patients. Poor control of patient positioning during set up, with common factors such as flexion and rotational alterations, frequently influence the apparent alignment on the lower limb radiograph [KPG90]. In addition, the image may be distorted by inclined x-ray beams [CSB<sup>+</sup>91, SCB<sup>+</sup>91]. Moreover, assessment of images which are down scaled in size to fit the screens loose precision and the two dimensional imaging of a three dimensional body part bears inaccuracy in itself.

Observer variability is a further factor possibly reducing the accuracy of manual measurements of limb alignment. Until recently, quantification of the knee alignment offered difficulties as a result of unacceptable reproducibility due to poor definition of landmarks and overall imprecise measurement techniques. A novel digital method using Image J software to designate landmarks digitally and to determine the angles formed by the femoral and tibial axes was presented in [GB07]. Using this method, the limits of agreement between duplicate assessments performed on separate days were reported to be  $+0.43^\circ$  to  $-0.37^\circ$  resulting in a minimally detectable change of  $0.4^\circ$  as compared to repeated manual measurements, the limits of agreement of which were stated to be  $+1.65^\circ$  to  $-1.55^\circ$ , yielding a minimal detectable change of approximately  $1.6^\circ$ .

Alternative measures are the anatomic-axis angle measured on a posteroanterior fixed-flexion knee radiograph or on an anteroposterior full-limb radiograph, which are correlated with the angle measured on the full-limb radiograph and have the potential to provide useful information regarding the risk of progression of knee OA when a full-limb radiograph is not available [KVWM05].

Computer-assisted angle measurement on digital total-leg radiographs offers

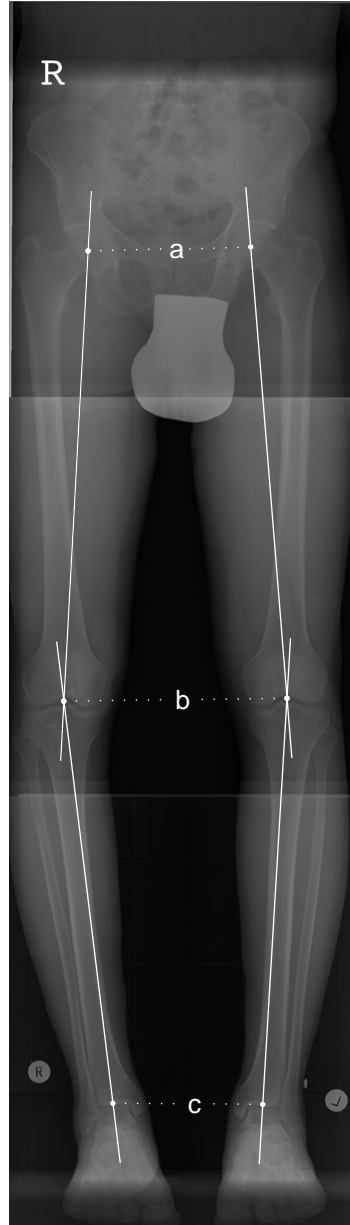


Figure 1.1: Full leg radiograph and definition of the knee alignment angle: the mechanical axis angle is defined by three points: the center of the femoral head (a), the center of the knee (b) and the center of the ankle mortise (c)

a significantly lower evaluation time while showing no significant angle differences compared to conventional radiographic systems [SSP<sup>+</sup>04]. So far, however, no fully automatic measuring method exists.

## 1.4 The aim of this thesis

The aim of this thesis is to develop a fully automatic method for assessing the knee alignment angles on full-limb radiographs. The automatic measurement should agree with the results of conventional assessment to an extent, so that it can be considered suitable for replacing manual methods. The method has to yield repeatable results to allow for the detection of even little change. Computer-aided manual assessment will serve as standard of reference. The automatic method is to bear comparison with respect to variability and precision.

The development of an automatic measurement method meeting these requirements involves selection and enhancement of suitable methods for

- automatic bone localization,
- accurate and robust shape delineation,
- coping with partially missing information (cropped shapes) and
- coping with overlapping and cluttered structures

## Chapter 2

# Automatic image interpretation

Many approaches have been proposed for recognizing objects and locating structures in medical images, which can be subdivided into data-driven methods and model-based methods [Coo04]. *Data-driven approaches* use only the image data available and examine it at a low level, looking for local structures such as edges or regions. Then, these low-level features are assembled into groups in an attempt to identify objects of interest. But without a global model of the expected appearance and shape of the object, this approach is difficult, prone to failure and usually requires considerable user-interaction during the process. *Model-based approaches* use a model of an object of interest to support its detection in an image. Prior knowledge contained in the model can be used to resolve the potential confusion caused by structural complexity, provide tolerance to noisy or missing data, and provide a means of labeling the recovered structures. The knowledge of the expected shapes of structures, their spatial relationships, and their grey-level appearance can be used to restrict the search to “plausible” interpretations. However, in the interpretation of highly complex and variable structures, typically the case in medical images, it has proved to be problematic to develop specific models, which are capable of representing only “legal” examples of the modeled object, while allowing for natural variability. This is mainly due to the required amount of training data necessary to achieve sufficient generalization performance.

*Deformable models* maintain the essential characteristics of the class of objects they represent, but can deform to fit a range of examples. In order to obtain specific models of variable objects, it is necessary to acquire knowledge of how they vary. Where structures vary in shape or texture, it is possible to learn to distinguish between plausible and illegal variations. Statistical models can capture specific patterns of variability in shape and gray-level appearance [CT92, CET98].

Kass et al [KWT88] introduced energy minimizing curves called *Active Contour Models* or “snakes”. An external energy term drives the curve toward image features, while an internal energy term aims to impose smoothness on the curve. Although they are useful for locating the outline of general amorphous objects,

they are not optimal for locating objects of a known shape, since they impose no model other than smoothness.

*Active Shape Models* [CTCG92] are based on statistical models and belong to the class of deformable models. They use a flexible model, which is based on the distribution of boundary points in training shapes and is thus called Point Distribution Model (PDM). By constraining the deformation based on the learned statistical prior, the model will only deform to fit the data in ways consistent with the training set. Active Shape Models are discussed in more detail in Sect. 2.2.

*Active Appearance Models* [CET98] contain a statistical model of the shape and grey-level appearance of the object of interest. They can generalize to almost any structure and are even capable of synthesizing new instances of the target object. The Active Appearance Model is a generalization of the Active Shape Model approach. It uses not only the information near the modeled edges, but also all textural information in the image region covered by the target object.

A comprehensive survey of deformable models used in medical image analysis is given in [MT96].

The remainder of this chapter is structured as follows:

- Sect. 2.1 explains how the standard of reference data is obtained from a set of sample images.
- Sect. 2.2 describes how statistical models of shape and local appearance can be built from the samples.
- In Sect. 2.3 two alternative methods are presented, which calculate a coarse estimate of the bones locations. This estimate is used to initialize a subsequent search procedure.
- Sect. 2.4 explains the basic Active Shape Model search algorithm.
- In Sect. 2.5 several enhancements to the basic search algorithm are introduced, which lead to better result accuracy and robustness.
- Sect. 2.6 describes how the alignment angle is calculated, after the lower limb bones have been delineated automatically.
- Finally, Sect. 2.7 summarizes the complete angle measurement procedure.

## 2.1 Obtaining the standard of reference data

In this thesis, we aim at finding the mechanical axes of the lower limbs by segmenting the right and left femora and tibiae and identifying the three specific points  $p_1$ ,  $p_2$ ,  $p_3$  describing the alignment angle (see Fig. 1.1). For the automatic delineation of the bones *Active Shape Models* are used, which employ statistical shape models of the object of interest. In order to build these models, a set of examples (in the following referenced as training set) is needed. In detail, we need samples of all four lower limb bones, which are gathered from a number of

full-leg radiographs and represented by a finite set of points along their contours, called *landmarks*.

### 2.1.1 Landmarks

A landmark is a point of correspondence on each object that matches between and within populations ([Ste00]). An  $nD$ -Landmark is a point  $x \in R^n$  in an  $n$ -dimensional space. Since in this work we solely use 2D outlines of objects, we define a 2D-landmark as point  $x \in R^2$ . In the following, when the term landmark is used, it always refers to 2D-landmarks.

One can roughly distinguish between three types of landmarks [DM98, CTCG95]:

- **Anatomical landmarks**

Points assigned by an expert that correspond between organisms in some biologically meaningful way.

- **Mathematical landmarks**

Points located on an object according to some mathematical or geometrical property, i.e. high curvature or an extremum point.

- **Pseudolandmarks**

Constructed points on an object either around the outline or between landmarks.

The identification of corresponding model points along the contours is a prerequisite for building statistical models of shape. The models will capture legal variations in shape by describing the dependencies between the variations of individual landmark positions. If the model points do not correspond between the samples, i.e. if they do not represent the same particular part of the object boundary, the method will fail to capture shape variability.

Manually annotating a sufficiently large number of landmarks of an object on several example images is a tedious and time consuming task. It has to be executed by a human expert who is familiar with anatomy and experienced in interpreting the type of medical image in question. It is not sufficient to indicate the positions of anatomical landmarks, since in most cases there are not enough of such landmarks to give a dense definition of an object. Besides, due to intra- and inter-reader variabilities, the required accuracy is often difficult to attain.

To define boundary more precisely, the anatomical landmarks can be augmented by equally spaced intermediary points between them. However, in practice the manual annotation of corresponding landmarks on a series of images still requires a tremendous effort. In three dimensions it becomes almost impossible ([DTC<sup>+</sup>02]). Therefore, a variety of automatic and semi-automatic methods have been developed to assist in this task [BMP01, Boo96, HTB00, KG92, RCB97, RMP<sup>+</sup>96, SP95, SLH91, KNWT02]. An overview and timeline of the various approaches can be found in [Coo04].

Although individual landmarks are hard to mark with sufficiently high precision, there exist well-established methods to semi-automatically obtain continu-



ous delineations of object contours. The approach taken in this thesis therefore involves two steps:

1. the continuous outlines of the objects of interest are delineated manually in a sample image, assisted by a method described in the following section.
2. landmark correspondences on these contours are established by an automatic method proposed in [DTC<sup>+</sup>02], which will be presented in Sect. 2.2.1.5.

### 2.1.2 Delineating the bones with Live-wire

The *Live-wire algorithm*, introduced in [MMBU92] and [USB92], provides a tool for efficient, accurate, and reproducible boundary extraction which requires only minimal user input. An objects boundaries can be extracted in one-fifth of the time required for manual tracing, with 4.4 times higher accuracy and 4.8 times higher reproducibility [BM97].

Boundary detection is formulated as a graph searching problem. The nodes of the graph represent pixels in the image and the edges are created in a 8-pixel neighbourhood. The user specifies a start node (called "seed point") manually from which the algorithm searches for an optimal path to a goal node, specified by the current position of the mouse pointer. When the mouse position is moved in proximity to an object edge, a "live-wire" boundary snaps to, and wraps around the object of interest (see figure 2.1). Selection of a new seed point freezes the selected boundary segment, and the process is repeated until the boundary is complete.

The optimal path or boundary is defined as the minimum cumulative cost path from a start node to a goal node, where the cumulative cost of a path is the sum of the local costs (or edge link) on the path. The local costs are calculated as a weighted sum of different component cost functionals, which exhibit the desired property to be low at positions of strong edges. These functionals include

- laplacian zero-crossing,  $f_Z$
- gradient magnitude,  $f_G$  and
- gradient direction,  $f_D$ .

Thus, we can write the local costs  $l(p, q)$  of the path from pixel  $p$  to a neighbouring pixel  $q$  as

$$l(p, q) = \omega_Z f_Z(q) + \omega_G f_G(q) + \omega_D f_D(p, q) \quad (2.1)$$

where  $\omega_Z$ ,  $\omega_G$  and  $\omega_D$  are the respective weights assigned to the individual feature function, with empirical default values of  $\omega_Z = 0.43$ ,  $\omega_G = 0.43$  and  $\omega_D = 0.14$  [BM97].

The gradient magnitude,  $f_G$ , is used for first-order positioning of the live-wire, since it is a strong measure of edge strength. The function is defined as

$$1 - \frac{G}{\max(G)} \quad (2.2)$$

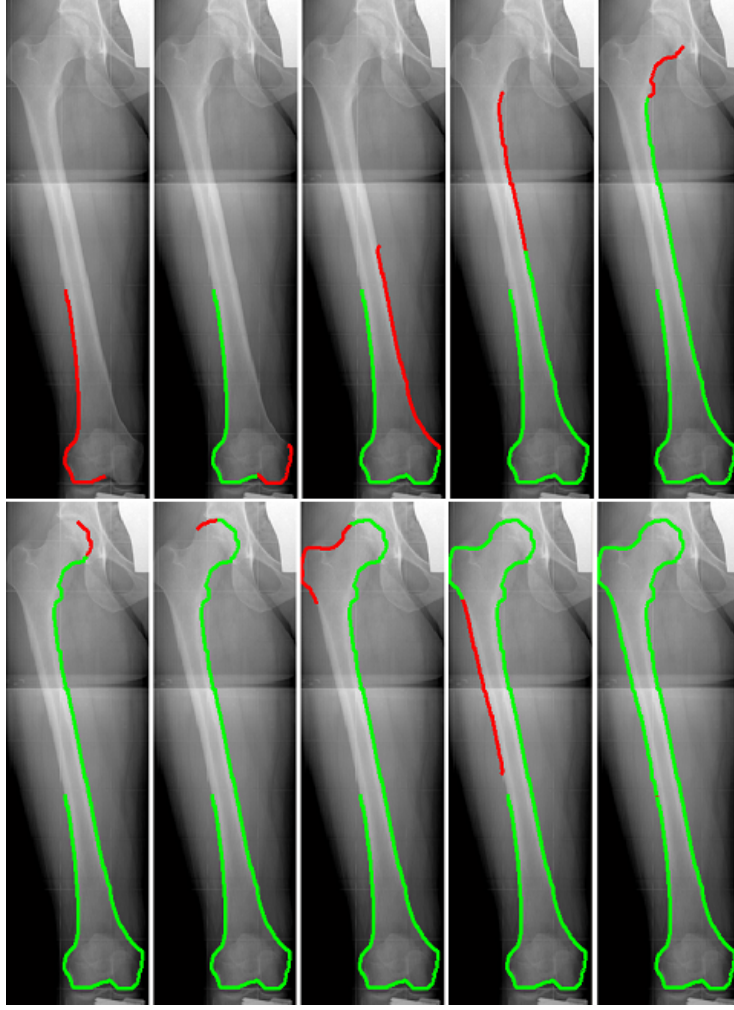


Figure 2.1: The contours of the right femoral bone captured with Live-Wire. Only ten points have to be selected manually to segment the complete outline of the bone.

to response with low values to high image gradients. The laplacian zero-crossing,  $f_Z$ , provides a second-order fine-tuning of the final boundary position and is defined  $f_Z(q) = 0$ , where the convolution of the image with a laplacian edge operator  $I_L(q) = 0$  or has a neighbor with a different sign, and  $f_Z(q) = 1$  otherwise.

The third component functional, gradient direction, restricts sharp changes in boundary orientation. Let

$$G(p) = (g_x(p), g_y(p)) \quad (2.3)$$

be the image gradient at point  $p$  with its components in the x and y direction,

respectively, and let

$$D(p) = \frac{(g_y(p), -g_x(p))}{\|G(p)\|} \quad (2.4)$$

(i.e. the unit vector normal to the gradient direction at point  $p$ ) then the gradient direction cost is defined

$$f_D(p, q) = \frac{2}{3\pi} (\cos[d_p(p, q)]^{-1} + \cos[d_q(p, q)]^{-1}) \quad (2.5)$$

where

$$d_p(p, q) = D(p) \cdot L(p, q) \quad d_q(p, q) = L(p, q) \cdot D(q) \quad (2.6)$$

are vector dot products and

$$L(p, q) = \begin{cases} q - p & \text{if } D(p) \cdot (q - p) \geq 0 \\ p - q & \text{if } D(p) \cdot (q - p) < 0 \end{cases} \quad (2.7)$$

is a unit vector representing the bidirectional link between pixels  $p$  and  $q$ . Thus,  $d_p(p, q)$  is the angle between the unit vectors  $D(p)$  and  $L(p, q)$ . The more the directions of image gradient and link between two pixels agree, the lower the cost that will be assigned to those pixels by the gradient direction feature  $f_D$ .

### 2.1.3 The training instances

The training instances were gathered from total-leg radiographs of 15 patients, 12 male and 3 female, with a mean age of 29.4+6.9 years. The spatial resolution of the image data was 0.3mm/pixel. The contours of the bones were captured manually by a third-year radiology resident (six months of musculoskeletal subspecialty training), aided by the live-wire algorithm described in the previous section. Two of the right tibiae and one right femoral bone had to be excluded as they were either cropped or occluded to a large part by plates and screws, so that their outlines could not be reconstructed. Hence, the models were built from a training set of 14 instances of the right femoral bone, 13 instances of the right tibial bone and 15 instances of both the left femoral bone and the left tibial bone.

## 2.2 Active Shape Models

Active Shape Models (ASMs), introduced in 1992 by Cootes et al. [CT92], are widely used in medical segmentation [LPBK07, PLW<sup>+</sup>07, TR01, CHTH94]. They provide a means to locate structures in images by utilizing prior knowledge about shape and how it varies as well as local texture along the contour boundaries. This enables to constrain the search to only plausible shapes. The prior knowledge is acquired during an initial training phase, where several manually annotated instances of the object of interest are presented to a training algorithm, which builds statistical models of shape and local texture. The search procedure involves alternating steps of local texture matching and global shape subspace constraint. The first searches within a limited range to move the contour points to a position that optimally fits the texture model, while the latter limits the movements to maintain shape plausibility. The model leads to a

reliable delineation, and can cope with a certain amount of ambiguous image content like overlapping structures in radiographs, or low contrast.

Although ASMs can be applied to shapes in 3D or even higher dimensional space, we only consider the 2D outlines of the objects we are looking for. Furthermore, we restrict the notion of contours to finite sets of contour points, which we call 2D-landmarks or just landmarks in the following.

## 2.2.1 Modelling shape

Sect. 2.1.2 described a way to collect object outlines from training images. In practice, the outlines are defined by a large but finite set of boundary points. However, theoretically the contours are continuous, i.e. they are defined by infinite sets of points. From these sets of boundary points, we need to select model points or *landmarks*. As stated earlier, landmarks represent key points of correspondence, which allow consistent localization on different images.

In section 2.2.1.5, a method is presented, which is able to select the corresponding model points automatically from a given set of sample outlines. This approach involves building shape models and optimizing an objective function to find optimal correspondences. Therefore, we will first discuss shape modelling and will present this method afterwards.

### 2.2.1.1 Shape and Pose

The training set we assembled consists of example outlines, which all are similar in shape but have different absolute positions in the respective images. Moreover, they are also differently scaled and rotated. In the following, we will distinguish between the notions of *shape* and *pose*, according to the following definition:

- **Shape** Given an  $n \times m$ -matrix  $\mathbf{A}$ , the column vectors  $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$  of which denote points in  $R^n$ , then shape is defined as the equivalence class with respect to the relation

$$\mathbf{A} \sim \mathbf{B} : \Leftrightarrow \exists \mathbf{T} : \mathbf{B} = \mathbf{T}^{-1} \mathbf{A} \mathbf{T} \quad (2.8)$$

where  $\mathbf{T}$  is the similarity transform matrix. In other words, the shape is all the geometrical information that is invariant to location, scale and rotation (cf. [DM98]).

- **Pose** The pose of an object is the geometrical information about its position, scale and rotation w.r.t. some coordinate reference.

Since we want to model *shape*, all information about position, scale and rotation is neutralized by aligning the training set in a way so that the outlines, now represented by vectors of landmarks, correspond as closely as possible, which is achieved by a form of Generalized Procrustes Analysis. This step is described in the following section.

### 2.2.1.2 Alignment of the training set

In order to extract information about shape, the training instances have to be freed from the effects of scale, translation and rotation. This is achieved by aligning the samples to a common coordinate reference system. All instances of a certain shape are thereby transformed into the same representative of the shape equivalence class. More important, outlines of *similar* shape become similar w.r.t. the distances between corresponding points. This is a necessary precondition for deriving meaningful statistics. The required alignment is approached by a form of *Generalized Procrustes Analysis* (GPA)[Gow75], which was described in [CTCG92].

Let  $\mathbf{x}$  be a vector of  $n$  landmarks describing an instance of the training set,

$$\mathbf{x} = (\mathbf{x}_0, \mathbf{y}_0, \mathbf{x}_1, \mathbf{y}_1, \dots, \mathbf{x}_n, \mathbf{y}_n) \quad (2.9)$$

Let  $M(s, \theta)[\mathbf{x}]$  be a rotation by  $\theta$  and a scaling by  $s$ . To align two similar shapes  $\mathbf{x}_i$  and  $\mathbf{x}_j$  we are looking for a  $\theta$ ,  $s$  and a translation  $\mathbf{t} = (\mathbf{t}_x, \mathbf{t}_y)$  which minimize the weighted sum

$$E = (\mathbf{x}_i - \mathbf{M}(s, \theta)[\mathbf{x}_j] - \mathbf{t})^T \mathbf{W}(\mathbf{x}_i - \mathbf{M}(s, \theta)[\mathbf{x}_j] - \mathbf{t}) \quad (2.10)$$

where

$$M(s, \theta) \begin{bmatrix} x_{jk} \\ y_{jk} \end{bmatrix} = \begin{pmatrix} (s \cos \theta)x_{jk} - (s \sin \theta)y_{jk} \\ (s \sin \theta)x_{jk} + (s \cos \theta)y_{jk} \end{pmatrix} \quad (2.11)$$

and  $\mathbf{W}$  is a diagonal matrix of weights for each point. These weights are to modulate significance of individual points according to the stability of their position w.r.t. the other points in a shape.

In [CTCG92] the following definition of the weight matrix was suggested: let  $R_{kl}$  be the distance between landmarks  $k$  and  $l$  of a shape, and let  $V_{R_{kl}}$  be the variance in this distance over the set of shapes. The weight  $w_k$  for the  $k^{th}$  point can be chosen

$$w_k = \left( \sum_{l=0}^{n-1} V_{R_{kl}} \right)^{-1} \quad (2.12)$$

This definition assigns high weights to points with a low sum of variances, which means, that this point has a stable position with respect to the other points.

To align all shapes in a set the following iterative procedure is proposed in [CT04]:

1. Translate each example so that its centre of gravity is at the origin.
2. Choose one example as an initial estimate of the mean shape and scale so that  $|\bar{\mathbf{x}}| = 1$ .
3. Record the first estimate as  $|\bar{\mathbf{x}}_0|$  to define the default reference frame.
4. Align all the shapes with the current estimate of the mean shape.
5. Re-estimate mean from aligned shapes.
6. Apply constraints on the current estimate of the mean by aligning it with  $|\bar{\mathbf{x}}_0|$  and scaling so that  $|\bar{\mathbf{x}}| = 1$ .

7. If not converged, return to 4. (Convergence is declared if the estimate of the mean does not change significantly after an iteration)

This method aligns each shape so that the sum of distances of each shape to the mean  $D = \sum |\mathbf{x}_i - \bar{\mathbf{x}}|^2$  is minimized. The constraints applied to the current estimate of the mean in step 6 are necessary to re-normalize the mean. Without this, the algorithm would be ill-conditioned and wouldn't converge [CTCG92].

### 2.2.1.3 Principal Component Analysis

The *Principal Component Analysis* (PCA) or discrete Karhunen-Loève transform is a method of dimensionality reduction introduced in 1933 by Harold Hotelling [Ste00]. The original variable space is rotated such that variance along the coordinate axes is maximized. Then the axes are ordered according to their variances. The first  $n$  principal axes will span a subspace that can explain a desired amount of the variance in the original samples. By discarding all further axes, the high dimensional samples can be compressed and compactly represented in the low-dimensional principal subspace. Thereby, a number of correlated variables is transformed into a smaller number of uncorrelated variables called *principal components*.

Figure 2.2 illustrates this for a two-dimensional input space spanned by the axis-vectors  $\mathbf{x}$  and  $\mathbf{y}$ . The two *principal components*  $\mathbf{p}_1$  and  $\mathbf{p}_2$  build the new basis. Every data point can be described by a projection on the first component  $\mathbf{p}_1$ , i.e. a coordinate with respect to  $\mathbf{p}_1$ , preserving a maximum of data variance in the resulting description.

To represent a shape of  $n$  points, the  $(x, y)$  coordinates of the landmarks can be concatenated to

$$\mathbf{x}_i = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n) \quad (2.13)$$

so that the shape can be considered as a point in  $2n$ -dimensional space. The training set of shapes is assumed to constitute an ellipsoid structure [Ste00] the centroid of which can be estimated [CET98]:

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \quad (2.14)$$

The covariance matrix of the set can be given

$$\Sigma = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (2.15)$$

and by normalizing with the variance

$$V = \text{diag}\left(\frac{1}{\sqrt{\text{diag}(\Sigma)}}\right) = \begin{bmatrix} \frac{1}{\sigma_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sigma_n} \end{bmatrix} \quad (2.16)$$

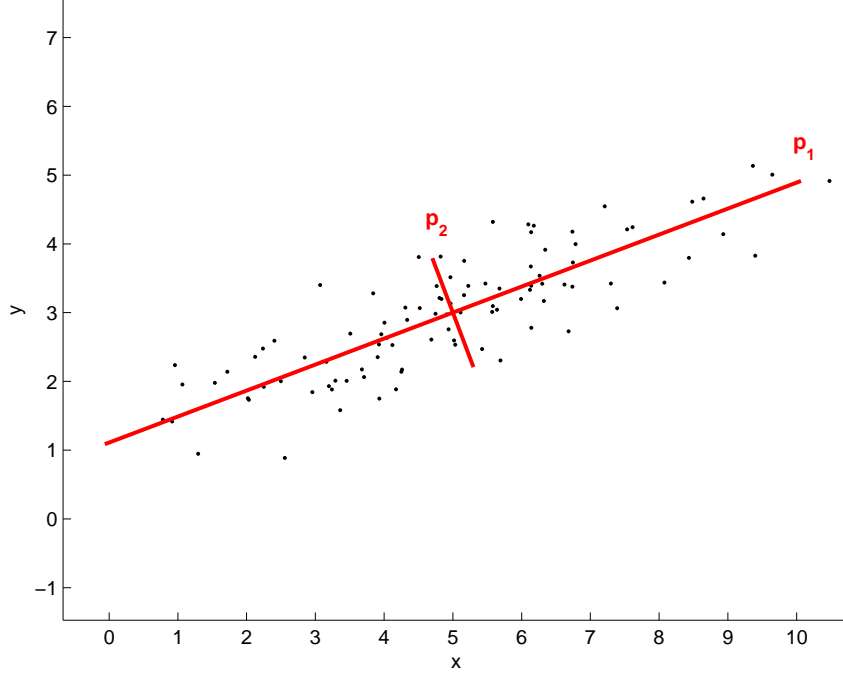


Figure 2.2: The principal axes of a two-dimensional data set.

the *correlation matrix* can be calculated [Ste00]:

$$\Gamma = \mathbf{V}\Sigma\mathbf{V}^T \quad (2.17)$$

This matrix denotes the strength of relationship between the point coordinates. Figure 2.4 illustrates the correlation matrix of the training set of 13 right femoral bones.

The principal axes are defined as the eigenvectors  $\Phi$  (or also called *modes*) of the covariance matrix. Their corresponding eigenvalues  $\lambda$  depict the variances of the data in the direction given by the respective eigenvector.

Every shape instance can be described by adding a linear combination of the eigenvectors to the mean shape  $\bar{\mathbf{x}}$ :

$$\mathbf{x} = \bar{\mathbf{x}} + \Phi\mathbf{b} \quad (2.18)$$

where  $\mathbf{b}$  is the vector of *shape parameters*. Note that the components of  $\mathbf{b}$  are ordered by decreasing amount of variation they account for. They denote the coordinates with respect to the new coordinate system built by the principal axes with the mean shape  $\bar{\mathbf{x}}$  being the origin. Figure 2.5 illustrates deformations of the mean shape of the right femoral bone using the first three principal modes.

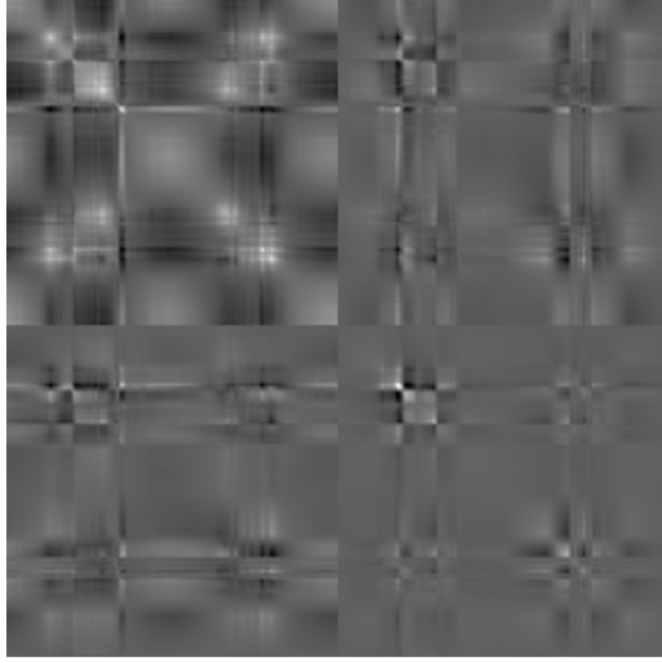


Figure 2.3: Covariance matrix of 13 right femoral bones with 256 landmarks each.

#### 2.2.1.4 Statistical Model of Shape Variation

The previous sections described how a training set of example outlines can be obtained from a set of sample images and how the instances can be aligned to neutralize pose information. After these steps, the remaining variations in the model point's positions reflect deformations of the shape. Applying PCA to the set of the  $n_s$  resulting shape representations with  $n_p$  model points each, yields a model with a mean shape  $\bar{\mathbf{x}}$  and a set of  $m = \min(2n_p, n_s)$  modes of variation  $\Phi = \{\phi_1, \dots, \phi_m\}$ , which are ordered by decreasing eigenvalues  $\lambda_i$ , i.e.  $\forall i = 1, \dots, m-1 : \lambda_i > \lambda_{i+1}$ . An instance of the shape model can be constructed by

$$\mathbf{x} = \bar{\mathbf{x}} + \Phi \mathbf{b} \quad (2.19)$$

where the parameter vector  $\mathbf{b}$  is built by the coefficients of  $\mathbf{x} - \bar{\mathbf{x}}$  with respect to  $\Phi$ .

To avoid overfitting the model to the training data and actually modeling noise, only the first  $t$  modes are chosen for the model. Since the variance along the direction given by the eigenvector  $\phi_i$  is determined by the corresponding eigenvalue  $\lambda_i$ , it is straightforward to retain a fraction  $p_v$  of the variation by



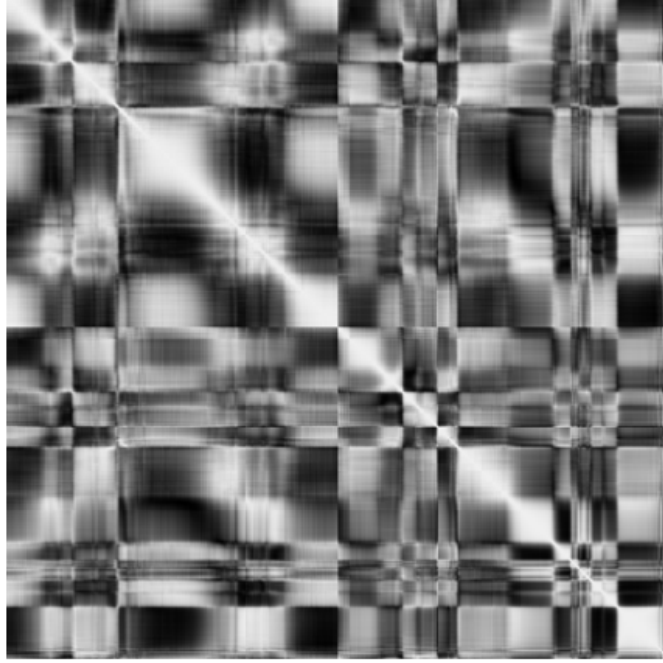


Figure 2.4: Correlation matrix of 13 right femoral bones with 256 landmarks each.

discarding all eigenvectors  $\phi_{\mathbf{t}}$  with

$$\sum_{i=1}^{t-1} \lambda_i \geq p_v \sum_{i=1}^m \lambda_i. \quad (2.20)$$

For this application, a number  $n_p = 256$  of model points was chosen for each bone, so that the original shape space had a dimensionality of  $2 \times 256 = 512$ . Choosing a fraction  $p_v = 0.95$ , the right femoral bone can be modeled using the first 13 modes. Thus, an instance of this bone will now be described by a 13-dimensional parameter vector. For the right tibia 12 modes were necessary, and the left femoral and tibial bones were modeled using 14 modes each. To illustrate the decay of the eigenvalues, figure 2.6 shows a bar graph of the eigenvalues of the training set of right femoral bones.

The remaining modes of variation  $\Phi$  determine the modeled directions of deformation learned from the training set. But not all deformations of the mean shape along these directions yield an instance which is similar to those in the original training set. If we wish to generate only plausible shapes, the shape parameters  $\mathbf{b}$  must be chosen from the distribution  $p(\mathbf{b})$ , which has to be estimated from the training examples, such that  $p(\mathbf{b}) \geq p_t$ , where  $p_t$  is some threshold on the probability density function. This threshold should be chosen

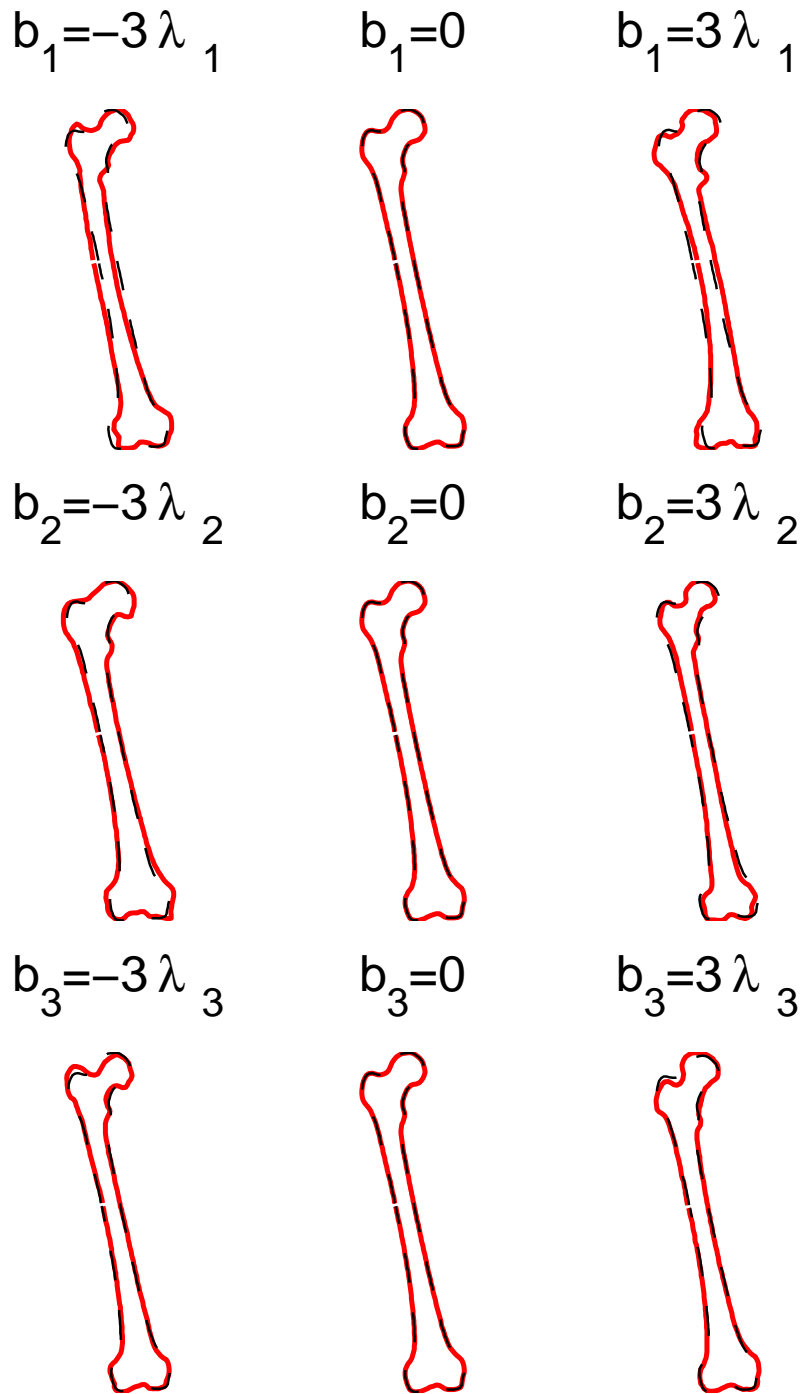


Figure 2.5: Deformation of the mean shape using the first three modes

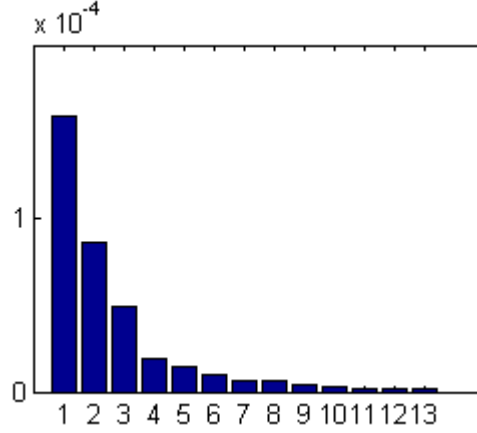


Figure 2.6: The decrease of eigenvalues of the training set of right femoral bones

so that a high percentage of the training set passes it. For this application the distribution can be approximated as a gaussian, since the changes in viewing positions are too small to cause non-linear shape variations. If we further assume that  $b_i$  are independent, then

$$\log p(\mathbf{b}) = -0.5 \sum_{i=1}^t \frac{\mathbf{b}_i^2}{\lambda_i} + c, \quad (2.21)$$

where  $c$  is constant. Limiting each element  $b_i$  of the shape parameters can constrain the deformations to only plausible shapes. Choosing

$$|b_i| \leq 3\sqrt{\lambda_i} \quad (2.22)$$

will let about 99.7% of the training instances pass (still assuming gaussian distribution). Alternatively,  $\mathbf{b}$  can be constrained to be in a hyperellipsoid,

$$\left( \sum_{i=1}^t \frac{b_i^2}{\lambda_i} \right) \leq M_t, \quad (2.23)$$

where  $M_t$  is chosen using the  $\chi^2$  distribution [CT04].

### 2.2.1.5 Automatically establishing landmark correspondences with the Minimum Description Length Approach

Identification of corresponding landmarks is a prerequisite for building statistical models of shape. Manual annotation is often infeasible, due to the large number of model points needed to give a dense definition of an object.

A fully automatic method for solving the correspondence problem to build statistical shape models from a training set of example boundaries (or surfaces) was proposed in [DTC<sup>+</sup>02] (earlier versions of this approach were published in

[KT98] and [DCT01]).

Dense correspondence over a set of training boundaries is established by defining a *parameterization* for each instance of the training set.

**Parameterization** A parameterization  $\psi_i(t)$  maps the indices  $t$  of the boundary points of an instance  $S_i$  of a shape to the indices of the boundary points of some reference instance, e.g. the *mean shape*  $\bar{S}$ . In other words, the parameterization governs the spacing between the points that are chosen to sample the shape. "Good" parameterizations satisfy the requirement, that for all indices  $t$  the point with the index  $\psi_i(t)$  of instance  $S_i$  corresponds to the  $t^{th}$  point of the reference instance.

Finding the correct parameterizations  $\{\psi_i\}$  of the training shape boundaries is treated as an explicit optimization problem. The basic idea is to find the parameterizations of the training shapes that yield, in some sense, the "best" model. Kotcheff and Taylor [KT98] describe an approach in which the best model is defined in terms of "compactness", as measured by the determinant of its covariance matrix. In [DTC<sup>+</sup>02] the objective function is defined in an information theoretic framework, with a rigorous theoretical basis. The quality of a model is measured by how efficiently the entire training set can be described, i.e. a *minimum description length* (MDL) [Ris83] criterion is adopted.

If the description had to be transmitted to some receiver, the message would have to include both the encoded model parameter values and the encoded data values. A poor quality of fit between the model and the data will increase the coding length of the data. A more complex model may allow for better compression of the data, but the costs for transmitting the model parameters will rise. Minimizing the overall description length thus means balancing the model's complexity against its fitness w.r.t. the data. Here, the description length is used as an objective function for optimization of correspondence between the shapes. The better the parameterizations of the individual shapes, the shorter the description length of the shape model parameters and the encoded shapes will be.

To construct a shape model we first need to change to a coordinate system whose axes are aligned with the principal axes of the training set. From the eigenvectors  $\{\phi_i\}$  of the covariance matrix we build the orthonormal set

$$\tilde{\phi}_m = \frac{\phi_m}{\|\phi_m\|}, m \in \{1, \dots, n_s - 1\} \quad (2.24)$$

where  $n_s$  is the number of training shapes. The new coordinates with respect to the principal axes can now be written

$$\mathbf{y}_m \equiv \mathbf{x} \cdot \tilde{\phi}_m \quad (2.25)$$

To describe this transformation the  $n_s - 1$   $n_p$ -D vectors  $\{\tilde{\phi}_m\}$  have to be transmitted. Their code length depends only on the number  $n_s$  of shapes and

the number  $n_p$  of points per shape and, hence, is constant for a given training set. The set of coordinates  $\mathbf{Y}_m = \{\mathbf{y}_{im}\}$ ,  $i \in \{1, \dots, n_s\}$  for each principal axis  $\tilde{\phi}_m$  is modeled using a 1-D Gaussian. To calculate their description length the data needs to be bounded and quantized. Given the original coordinates are strictly bounded:

$$\forall_{\alpha \in \{1, \dots, n_p\}, i \in \{1, \dots, n_s\}} : -\frac{r}{2} \leq x_{i\alpha} \leq \frac{r}{2} \quad (2.26)$$

then

$$R = r\sqrt{n_p} \quad (2.27)$$

with

$$\forall_{i \in \{1, \dots, n_s\}, m \in \{1, \dots, n_s-1\}} : -\frac{R}{2} \leq y_{im} \leq \frac{R}{2} \quad (2.28)$$

gives the upper-bound for the new coordinates  $\{y_i\}$  [DTC<sup>+</sup>02]. The quantization parameter  $\Delta$  is determined by comparing the original shape and the quantized shape. For boundaries from pixellated images  $\Delta$  is typically in the order of pixel size. The quantization parameter will also give a lower bound for the modeled variance:

$$\sigma_{min} = 2\Delta. \quad (2.29)$$

An expression for the description length of one-dimensional, bounded, and quantized data, coded using a Gaussian model was derived in [DTC<sup>+</sup>02]:

$$D_m = \begin{cases} \ln\left(\frac{R}{\Delta}\right) + D^{(1)}(\hat{\mathbf{Y}}_m, R, \Delta) & \text{if } \sigma_m \geq \sigma_{min} \\ \ln\left(\frac{R}{\Delta}\right) + D^{(2)}(\hat{\mathbf{Y}}_m, R, \Delta) & \text{if } \sigma_m < \sigma_{min} \\ \ln\left(\frac{R}{\Delta}\right) & \text{else} \end{cases} \quad (2.30)$$

where  $\hat{\mathbf{Y}}_m = \{\hat{y}_i\}$  denotes the quantized values of the original continuum coordinates  $\mathbf{Y}_m = \{y_i\}$ .

Letting  $n_g$  be the number of directions which satisfy the first of the above criteria and  $n_{min}$  the number of directions satisfying the second, then the total description length of the training set can be written

$$D = (n_s - 1) \ln\left(\frac{R}{\Delta}\right) + \sum_{p=1}^{n_g} D^{(1)}(\hat{\mathbf{Y}}_p, R, \Delta) + \sum_{q=n_g+1}^{n_g+n_{min}} D^{(2)}(\hat{\mathbf{Y}}_q, R, \Delta) \quad (2.31)$$

Since the first term is constant for a given training set the objective function  $F$  for the optimization task can be defined

$$F = \sum_{p=1}^{n_g} D^{(1)}(\hat{\mathbf{Y}}_p, R, \Delta) + \sum_{q=n_g+1}^{n_g+n_{min}} D^{(2)}(\hat{\mathbf{Y}}_q, R, \Delta) \quad (2.32)$$

As stated earlier, the correspondence problem is cast as one of finding the best parametrization  $\psi_i$  for each shape. Hence,  $F$  is minimized by manipulating  $\{\psi_i\}$ . By demanding the parametrization  $\psi_i(t)$  of each shape to be a monotonically increasing function of  $t$ , the point ordering and, consequently, the correspondences can be constrained. In two dimensions, it is sufficient to enforce the ordering of points on the boundaries according to arc-length. However, no such ordering exists on surfaces. To overcome this, a piecewise-linear representation of the parametrization is suggested in [DTC<sup>+</sup>02], which enforces an implicit ordering of the nodes and guarantees a diffeomorphic mapping even for surfaces in three dimensions. For a single shape, the parametrization is constructed following the below steps:

1. Chose an initial node from the shape outline. This node will represent the initial set of *parent nodes* and count for both the *origin* and the *endpoint*.
2. Create a child node between each adjacent pair of parent nodes. The child node will be described by the fractional distance along the curve between it's parent nodes.
3. Form a new set of initial nodes from the current parent nodes and the new child nodes and enter the next level of recursion (Step 2), until a sufficient number of nodes has been created.

In this fashion, a parametrization is described by a set of fractional distances defining the positions of the child nodes relative to their respective parents. The position values range from 0 to 1, where 0 means "on the left neighbour", 1 stands for "on the right neighbour" and 0.5 denotes the position in the center between the two neighbours. The number of nodes determines the degree of refinement and depends, in general, on the complexity of the training shapes.

From the parametrizations  $\psi_i(t)$  constructed this way, an arbitrary number of corresponding points can be obtained by sampling the shapes at equally spaced intervals of  $\psi_i(t)$ . The sampled shapes can then be evaluated by calculating the objective function  $F$ .

The optimization procedure can be sketched as follows:

1. Generate a parametrization for each shape to the same level of recursion
2. Sample the shapes according to their parametrization
3. Build a model
4. Calculate  $F$
5. Vary parametrization of each shape until  $F$  is optimal.

As  $F$  behaves highly non-linear and contains many local minima, a stochastic optimization method, like e.g. a Genetic Algorithm search, is preferable [DTC<sup>+</sup>02]. It is also important to note, that  $F$  can be minimized by collapsing all points on all shapes to a single part of the boundary. This undesirable result can be avoided by introducing a reference shape with a fixed parametrization.

To demonstrate the benefit of using the Minimum Description Length method, two models were built from a training set of outlines of the left femoral

bone. The first model was built using arc-length parametrization, while for the latter a parametrization optimized by applying the above described Minimum Description Length method was used. Figure 2.7 shows the first three modes of the resulting models when using different parametrizations. The parametrization that minimizes the description length yields a model that can only represent valid shape instances, while the model built with arc-length parametrization fails to capture valid shape variation correctly.

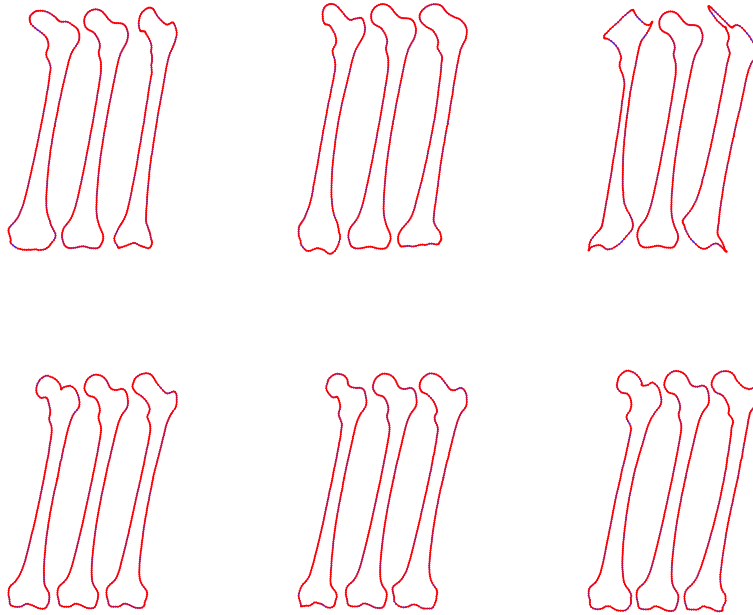


Figure 2.7: The first three modes of variation of a model built from the training set of outlines of the left femoral bone with arc-length parametrization (top row) and with a parametrization generated using the MDL method (bottom row). The model built with arc-length parametrization can represent invalid shape instances, whereas the model built with MDL parametrization can only represent valid shape instances.

### 2.2.2 Modelling local texture

In the previous chapter we focused on how the shape of an object and its “legal” variations can be modeled. If we want to use our shape model for image interpretation, we need a *drive* that governs the way the landmark positions are updated in order to fit the model to the target image. If we know that the model boundary always corresponds to an edge, we can look along profiles perpendicular to the current boundary through each model point and locate the strongest edge. But in practice, model points are often placed on weaker secondary edges or some other structure [CT01]. In this section, we therefore address a way to

learn a model of local image content in proximity of the landmarks from the training set.

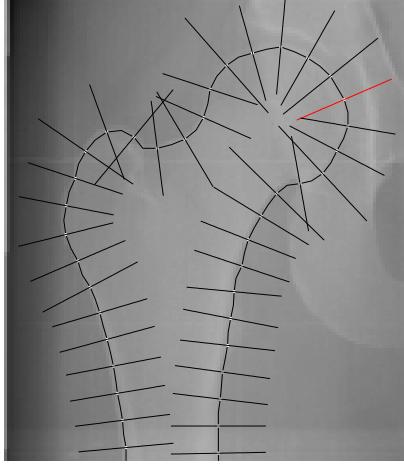


Figure 2.8: Gray value profiles orthogonal to the shape.

During training, the manual annotation of the object's outlines in each of the training images and the identification of a set of corresponding landmarks on the outlines of every object, provides the positions  $x_j^i \in \mathbb{R}^2$  of the landmarks. For each of these points we sample along a profile normal to the object boundary  $k$  pixels on either side of the point, as depicted in figure 2.8. The resulting  $k + 1$  pixels build a vector  $\mathbf{g}_j^i$  (Fig. 2.2.2). Global changes in illumination and contrast are largely eliminated by calculating the first derivative  $\mathbf{g}_j^i$  and normalizing it by dividing through by the sum of absolute element values (Fig.

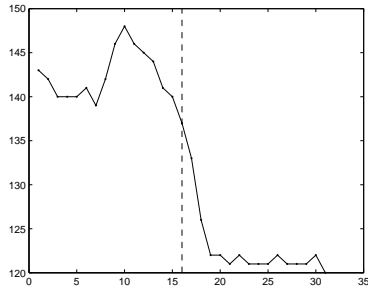


Figure 2.9: Gray value samples  $g_j^i$  along the profile.

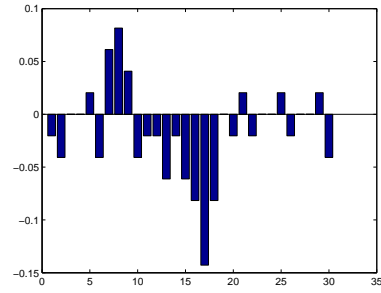


Figure 2.10: Normalized derivative of gray value samples  $g_j^i$  along the profile.



2.10),

$$\mathbf{g}_j^i \rightarrow \frac{\mathbf{g}_j^i}{\sum_{s=1}^{2k+1} |\mathbf{g}_{j,s}^i|} \quad (2.33)$$

The resulting set of texture vectors for each landmark  $j$  is assumed to be distributed as a multivariate Gaussian, and we estimate the mean  $\bar{\mathbf{g}}_j$  and covariance  $\mathbf{S}_j$  to obtain statistical model for the gray-level profile about the point [CT01].

### 2.2.3 Multi-resolution Active Shape Models

For this application a multi-resolution approach was used which incorporates a coarse-to-fine search strategy to increase both speed, robustness and quality of fit [CTL94]. First, the object is delineated in a coarse, low-resolution version of the image, and then it is refined on higher resolution versions. For this it is necessary to train a set of gray-level models for each landmark, one for every level  $L$  of a multi-resolution image pyramid. Level  $L = 0$  represents the original image. By smoothing and subsampling the image at level each level  $L < L_{max}$ , an image of level  $L + 1$  with half the number of pixels in each dimension is obtained. Hence, the models at the higher levels are coarser, but capture larger parts of the image. During search, the higher levels will allow large movements and will yield a coarse delineation of the target object, which is then used to initialize search on a finer level.

## 2.3 Coarse localization of the bones

Interpreting an image using a model is the task of finding the set of parameters, defining the shape and pose, which optimally match the image to the target image. Most generally, this is an optimization problem, where the optimum is defined by a fit function. The fit function is obtained by generating an instance of the model projected into the image and calculating a distance measure between this hypothesis and the image content. If we have no initial knowledge of where to look for the target object in an image, finding the best set of parameters to maximize the fit is a difficult optimization problem, which can, in principle, be solved with any general purpose optimizer, such as Genetic Algorithms or Simulated Annealing [HCT92]. However, given an initial approximation to the position of the target object in the image, local search techniques can be used to locate the optimum rapidly. Active Shape Models, discussed in Sect. 2.2, take advantage of the form of the fit function and provide one of such techniques [CT04].

Thus, a method for coarsely estimating the initial position of the target object in the image is needed. In this chapter, two possible approaches to this task are presented: an application specific approach (Sect. 2.3.1) and a more general approach based on sparse appearance models (Sect. 2.3.2).

### 2.3.1 An application-specific approach

In the course of this thesis, a simple application specific approach to roughly estimating position, scale and orientation of the femora and tibiae was developed. These estimates can be used for the initialization the segmentation algorithm for each respective bone.

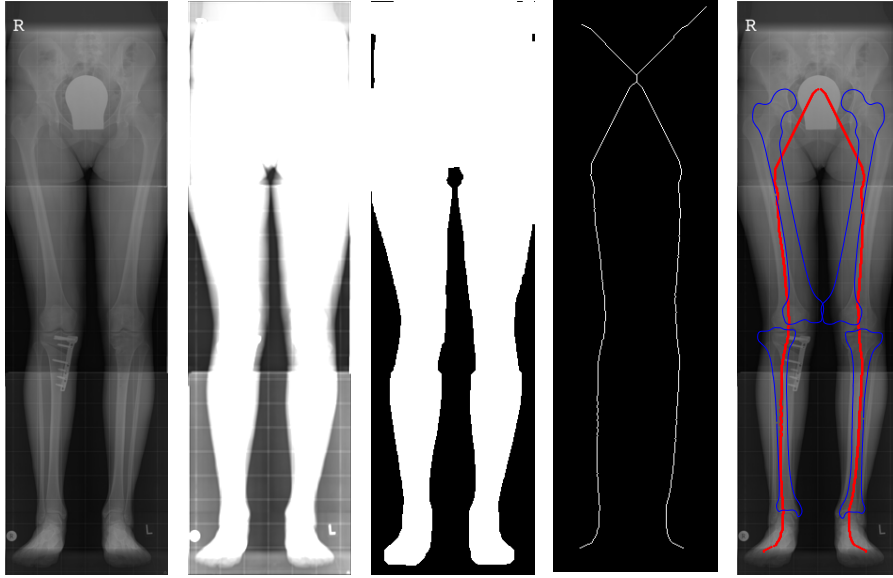


Figure 2.11: The lower limbs and hip are segmented from the background by filtering and thresholding. From the resulting shape the medial axis transform ("skeleton") is calculated by morphological thinning. The center of the hip is estimated by the position where the skeleton forks into two branches. The two branches themselves are used as approximation of the medial line through the bone shafts

As a first step, the image is down-scaled to 15% of its original size. To strengthen the contrast between foreground, i.e the patients legs and hip, and background, a filter  $\mathbf{F}$  is applied, which is calculated from a bivariate gaussian filter  $\mathbf{F}'$  by dividing through  $\max \mathbf{F}' = 1/(2\pi\sigma_x\sigma_y)$ . Empirically, choosing  $\sigma_x = 1$  and  $\sigma_y = 0.7$  has proved to yield expedient and stable results. By thresholding the filtered image, a binary image of the legs and hip is created. The threshold is chosen using *Otsu's method*, which minimizes the intraclass variance of the thresholded black and white pixels. Next, *morphological opening* is performed on the binary image in order to remove small objects and smooth the outlines of large objects. From the resulting shape the *medial axis transform* ("skeleton") is calculated by *morphological thinning*. Searching vertically downward in the image, the center of the hip is estimated by the position where the skeleton forks into two branches. The two branches themselves are used as approximation of the medial line through the bone shafts (figure 2.11).

### 2.3.2 Sparse Markov Random Field Appearance Models

Recently, Donner et al [DML<sup>+</sup>07] proposed a method based on Markov Random Fields (MRF) to detect anatomical structures by configurations of interest points. The approach incorporates the positions of the interest points as well as local features in their vicinity, and uses the information of all interest points to find an optimal mapping of the modeled object to the target image. The quality of fit is defined by the combined costs of non-rigid deformations and local descriptor differences. The MAX-SUM algorithm [Wer05] finds a solution that minimizes these costs in a single iteration.

In principle, arbitrary interest points and local descriptors can be used. For this application, symmetry-based interest points and local descriptors [DML<sup>+</sup>07] derived from Gradient Vector Flow (GVF) [XP98] were used, since the limbal bones exhibit a shape with a high degree of symmetry w.r.t. their anatomical axis. These interest points and local descriptors benefit from the ability of the GVF field to detect even weak structures while being robust to certain amounts of noise in the image.

The GVF field can be computed from an edge map or directly from a grey-level image and is given by a complex matrix  $\mathbf{G}$ . Maxima of the field magnitude  $|\mathbf{G}|$  mark areas of high image gradient and the start- and end-points of the field lines of  $\mathbf{G}$  are located at symmetry maxima. Thus, the symmetry interest points are defined by the local minima of  $|\mathbf{G}|$  (Fig. 2.12).

The orientation  $b_i \in [0, \pi]$  of the local region surrounding the interest point is calculated as

$$b_i = \angle \mathbf{G}(x_i + \Delta x_i, y_i + \Delta y_i), \quad (2.34)$$

which is the orientation of  $\mathbf{G}$  at a pixel in a local  $r \times r$ -pixel neighborhood satisfying

$$(\Delta x_i, \Delta y_i) = \underset{\substack{\Delta y_i \in \{-r/2, \dots, r/2\} \\ \Delta x_i \in \{0, \dots, r/2\}}}]{\operatorname{argmin}} \quad |(\angle \mathbf{G}(x_i + \Delta x_i, y_i + \Delta y_i) - \angle \mathbf{G}(x_i - \Delta x_i, y_i - \Delta y_i))|. \quad (2.35)$$

To estimate the scale  $s$  of the interest point, the mean distance to the two closest local maxima of  $|\mathbf{G}|$  in the direction of  $b_i \pm \pi$  is calculated.

The local descriptors are patches extracted from  $\mathbf{G}$  around the symmetry interest points, according to the scale  $s_i$  and orientation  $b_i$ , and re-sampled to a  $10 \times 10$  grid, as depicted in Fig. 2.13. Hence, the local descriptors encode information about the image gradients around the interest points, in a way invariant to scale and rotation of the object in the image.

The similarity between two descriptors can be measured using euclidean distance. However, as the orientation of an interest point is only defined up to  $\pm\pi$ , the actual distance between two descriptors  $\mathbf{D}_1$  and  $\mathbf{D}_2$  is calculated as

$$\min(\|abs(\mathbf{D}_1 - \mathbf{D}_2)\|, \|abs(\mathbf{D}_1 - \mathbf{D}_2^*)\|), \quad (2.36)$$

where  $\mathbf{D}_2^*$  denotes the second descriptor  $\mathbf{D}_2$  rotated by  $\pi$ .

A subset of the interest points on a model image is selected to represent the object of interest. The selected points build the nodes of a graph. The  $a$  edges are defined by the Delaunay triangulation of the points and represented by a

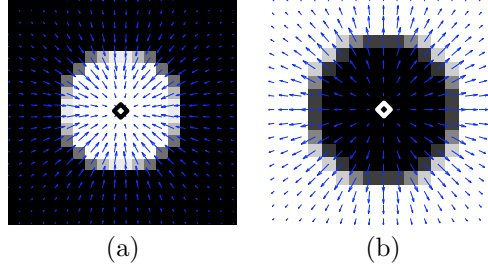


Figure 2.12: Examples of GVF with the detected symmetry interest points (diamonds). In the case of a symmetrical structure formed by a homogeneous region surrounded by a different gray level value the field points either towards (a) or away from (b) the local symmetry center of the structure. (Figure taken from [DML<sup>+</sup>07])

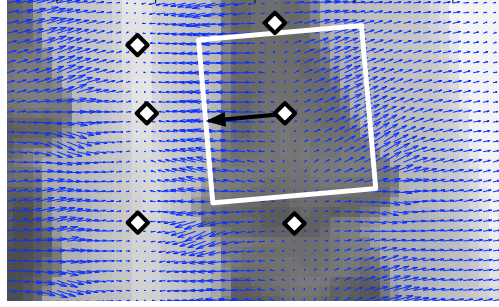


Figure 2.13: Descriptor extraction from the GVF field. Around each symmetry point patches are extracted from the vector field according to their scale and orientation. The patch is then resampled to a  $10 \times 10$  grid to form the actual descriptor. The image is displayed for better visualization, the symmetry points are marked as circles. (Figure and text taken from [DML<sup>+</sup>07])

set  $\mathcal{A}$  of index-tuples. To localize an object in a target image, all target interest points are considered as potential candidates to correspond to the model points. The  $N$  candidate correspondences of each model point are called *fields* or *labels* and the  $M$  model points *objects*. The Markov Random Field represents a graph with  $N^2$  edges fully connecting the labels of two adjacent objects, so that the total number of edges is  $aN^2$  (see Fig. 2.14). The labels as well as the edges have qualities (or weights) assigned to them. The quality of a label equals the negative distance between the local descriptors of the respective target interest point and model point. To calculate the quality of an edge between two labels  $n_i$  and  $n_j$ , its length and angle is compared to the edge  $\alpha$  between the corresponding objects. The edge quality is set to

$$e(\alpha, n_i, n_j) = -(|\text{length}(\mathcal{A}(\alpha)) - \text{length}(n_i, n_j)| + \gamma(|\angle(\alpha) - \angle(n_i, n_j)|)), \quad (2.37)$$

where  $\text{length}(h, k)$  represents the pixel distance between interest points  $k$  and  $h$ ,  $\angle(h, k)$  is the orientation of the edge and  $\gamma$  is a normalization factor to compensate for the different scale of angles and lengths.

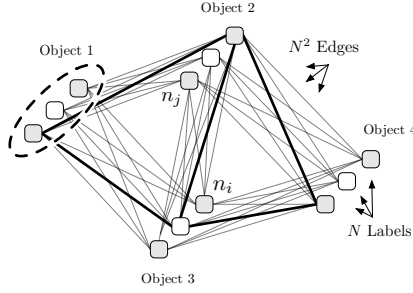


Figure 2.14: The MRF graph consists of  $M$  nodes (objects), each of which has  $N$  fields (labels). The labels of two adjacent nodes are fully connected by  $N^2$  edges. (Figure taken from [DML<sup>+</sup>07])

Selecting labels for each object, so that the sum of label and edge qualities of the resulting sub-graph becomes maximal, optimally fits the model to the target interest points. This can be achieved by the MAX-SUM solver.

The MAX-SUM (labeling) problem of the second order is defined as maximizing a sum of bivariate functions of discrete variables. The solution corresponds to finding a configuration of a Gibbs distribution with maximal probability, which is equivalent to finding a maximum posterior (MAP) configuration of a Markov Random Field with discrete variables [Wer05].

Letting  $\mathbf{C}$  be an  $M \times N$ -matrix, assigning qualities of fit between the  $N$  labels and the  $M$  objects, and  $\mathbf{E}$  an  $a \times N^2$ -matrix determining the  $a$  edge qualities for each of all  $N^2$  pairs of labels, the total quality of the label selection  $\mathcal{S} = \{n_1, \dots, n_M\}$  with  $n_i \in \{1, \dots, N\}$  is defined as

$$C(\mathcal{S}) = \sum_{m=1 \dots M} \mathbf{C}(m, \mathcal{S}(m)) + \sum_{\alpha=1 \dots a} \mathbf{E}(\alpha, \beta(E, \mathcal{S}, \alpha)), \quad (2.38)$$

where  $\beta(\mathbf{E}, \mathcal{S}, \alpha)$  denotes the column representing the quality of the edge between the labels chosen to represent the edge  $\mathcal{A}(\alpha)$ . The set of optimal labels is thus

$$\mathcal{S}^* = \underset{\mathcal{S}}{\operatorname{argmax}} C(\mathcal{S}). \quad (2.39)$$

The MAX-SUM approach doesn't solve the problem of a multi-label MRF exactly, as it is NP-hard. Still, if the graph is a tree the global optimum of Eq. (2.39) is guaranteed [Kol06], and otherwise, MAX-SUM takes various approximations into account to reach a possibly optimal solution.

Both, this method and the application specific approach described in Sect. 2.3.1, provide a coarse location estimate of the bone's positions.

## 2.4 Image interpretation with standard Active Shape Models

The generated models of shape and local texture can now be used to find new examples of the modeled object in images. In this section an iterative algorithm

is presented to find the shape and pose parameters, which cause the model to coincide with the structures of interest in the image.

Starting from a rough estimation of the instance  $\mathbf{X}$  of the model, the algorithm - using multi-resolution ASM - can be summarized as follows [CTL94, CTCG95]:

1. Set  $L = L_{max}$
2. While  $L > 0$ 
  - (a) Compute the model point positions of the current estimate of  $\mathbf{X}$  in the image at level  $L$ .
  - (b) Update the landmark positions according to image content.
  - (c) Transform these local deformations into adjustments to the pose, scale and shape parameters of the Point Distribution Model
  - (d) Enforce limits on the shape parameters to apply global shape constraints ensuring the shape of the model example remains similar to those of the training set.
  - (e) Return to (2a) unless more than  $p_{close}$  of the points are found close to the current position, or  $N_{max}$  iterations have been applied at this resolution.
  - (f) If  $L > 0$  then  $L \rightarrow (L - 1)$
3. The final result is given by the parameters after convergence at level 0.

This way, the models attempt to deform to better fit the data, but only in ways which are consistent with the shapes found in the training set. In the following, the particular steps are explained in detail.

#### 2.4.1 Updating the landmark positions according to image content

In order to fit the model instance  $\mathbf{X}$  to the image content, a region of the image around each model point is examined to determine a displacement which moves it to a better location. This is done by translating each landmark  $j$  along a line perpendicular to the shape and sampling  $m$  profile vectors  $\gamma_{j,d}, d \in \{1, \dots, m\}$  according to the computation of the gray-value vectors  $\mathbf{g}_j^i$  during training. The profile vector  $\gamma_{j,t}$ , which minimizes the *Mahalanobis distance* [Mah30] to the modeled texture  $\bar{\mathbf{g}}_j$  is chosen and landmark  $j$  is updated accordingly.

The Mahalanobis distance is defined

$$f(d) = (\gamma_{j,d} - \bar{\mathbf{g}}_j)^T \mathbf{S}_j^{-1} (\gamma_{j,d} - \bar{\mathbf{g}}_j), \quad (2.40)$$

where  $\mathbf{S}_j$  is the covariance of the gray value profile samples in the training set.

Repeating this for every model point gives a suggested new position for each point.

### 2.4.2 Updating pose and shape parameters

After having calculated new positions  $\mathbf{Y} = \mathbf{X} + d\mathbf{X}$  of all landmarks to improve the fit of the model instance to the image content, the shape model is fit to these new points. This is achieved by adjusting the shape and pose parameters in a way so that the resulting model points are as close as possible to the suggested positions  $\mathbf{Y}$ . Letting  $T_{X_t, Y_t, s, \theta}$  be a function, which performs a translation by  $(X_t, Y_t)$ , a rotation by  $\theta$  and scaling by  $s$ , the updated model instance  $\mathbf{X}'$  can be expressed

$$\mathbf{X}' = T_{X_t, Y_t, s, \theta}(\bar{\mathbf{x}} + \Phi \mathbf{b}). \quad (2.41)$$

The optimal fit can be found by minimizing the sum of squared distances

$$\|\mathbf{Y} - \mathbf{T}_{X_t, Y_t, s, \theta}(\bar{\mathbf{x}} + \Phi \mathbf{b})\| \quad (2.42)$$

between corresponding points in  $\mathbf{Y}$  and  $\mathbf{X}'$ . The search can be performed by the following iterative algorithm [CT04]:

1. Initialize the shape parameters,  $\mathbf{b}$ , to zero
2. Generate the model instance  $\mathbf{x} = \bar{\mathbf{x}} + \Phi \mathbf{b}$
3. Find the pose parameters  $(X_t, Y_t, s, \theta)$  which best map  $\mathbf{x}$  to  $\mathbf{Y}$
4. Invert the pose parameters and use to project  $\mathbf{Y}$  into the model co-ordinate frame:

$$\mathbf{y} = T_{X_t, Y_t, s, \theta}^{-1}(\mathbf{Y}) \quad (2.43)$$

5. Project  $\mathbf{y}$  into the tangent plane to  $\bar{\mathbf{x}}$  by scaling by  $1/(\mathbf{y} \cdot \bar{\mathbf{x}})$
6. Update the model parameters to match to  $\mathbf{y}$

$$\mathbf{b} = \Phi^T(\mathbf{y} - \bar{\mathbf{x}}) \quad (2.44)$$

7. Apply constraints on  $\mathbf{b}$ , so that  $p(\mathbf{b}) \geq p_t$ . See Eq. 2.21 and Eq. 2.22.
8. Go to step 2 until the search converges.

The search is considered converged, when applying an iteration produces no significant change in the pose or shape parameters. Usually, this approach converges in a few iterations.

## 2.5 Improvements to the ASM Approach

Applying Active Shape Models in their basic form already yields good results for some of the test instances (see Sec. 3). However, if the search is not initialized sufficiently well, the search procedure is prone to ending prematurely in (or close to) a local minimum. It can even pull the landmarks towards wrong directions. The application specific initialization approach may fall below the required accuracy, if the gauss-filtered gray-level intensity between the thighs drops insignificantly, and therefore a possibly too large fraction of the legs cannot

be separated. The estimated position of the center of the hip and, hence, the estimated positions of the limbal bones will be vertically displaced from their true locations. The initialization using Sparse MRF Appearance Models can fail if due to overlapping structures, partially missing data and changing intensities, a too little number of interest points can be identified to match the modeled landmarks.

But even if the initialization provides a considerable overlap with the bone in the target image, ASM-search may fail for full leg radiographs, since screws, plates, raster lines or the seam between two radiograph-tiles, that cannot be completely removed by filtering, cause an appearance which is only partially consistent with the learned model, resulting in a cost minimum that does not coincide with the correct match of the model. Moreover, the standard ASM segmentation result does not provide a measure of confidence, that would indicate such a failure to the clinical expert.

For this reason, we introduce an approach that continuously assesses the local *fit confidence* during an ASM search. We apply it to the delineation of lower limb bones in composite full-leg radiographs. The fit confidence contributes in two ways: 1. it allows for a transparent assessment of the reliability of the local delineation result, and 2. it can control the search procedure automatically and results in improved results for ambiguous and complex data. This measure allows to incorporate strategies for search method refinement, such as dynamic adaptation of search parameters, conditional inclusion of an alternative search step or re-initialization, which considerably improves accuracy and robustness of the results.

### 2.5.1 Related work

Several improvements to the ASM methodology have been proposed. Yan et al [YLL<sup>+</sup>03] proposed a method which effectively incorporates not only the shape prior and local appearance around each landmark, but also the global texture constraint over the shape, to perform more stable against varying initialization. They used *boosted regression* instead of the original eigen model profiles of the texture around each feature point, to learn the relationship between the displacement to the true feature location and the textural appearance of the local neighbourhood around each feature point.

Steward et al [SLT04] presented an uncertainty-driven hybrid of feature-based and intensity-based registration, where the alignment error variance is computed to generate registration constraints for the next iteration of the matching process. The search region is controlled by both the uncertainty in the current transformation estimate and the properties of the image locations to be matched.

Wang et al [WWL06] proposed and compared new local detectors, which essentially combined grey-level derivatives and edges sampled from an edge map to build improved point distribution models. They filled the sample vectors with edges and assigned polarity of the derivative or, by their preference, alternated edge and derivative to build a combination vector of double length. These combination vectors are then used for comparison instead of the normalized derivative vectors to improve the performance of the search procedure.



In a related line of work, Beichel et al. [BBL05] proposed robust active appearance models (AAM), that deal with occlusions and outliers during an AAM search, by discarding respective regions based on their coherence with the learnt model.

### 2.5.2 Estimating Fit Confidence

In this section, an approach to continuously assessing reliability throughout the search process is introduced. The reliability estimates can be used to control search and - if necessary - can caution the clinical expert about low confidence and possible failure. While the constant assessment of confidence improves the search performance, the assessment of the result confidence is an essential property for the application of automatic methods in a clinical setting.

During the search process local texture matching moves the landmarks to optimal positions within a certain search range. Ideally, in the last iteration the new landmark positions will accurately represent the outlines of the modeled object in the target image. In this case, the drive of local texture matching and shape constraint will agree perfectly. However, as long as the shape and pose parameters are still too far apart from their correct values, the new landmark positions will not form a plausible shape and will thus be corrected by the limitation of the shape parameters. As the search proceeds, the shape will either be moved out of the boundaries of the target image, or the landmark positions will engage in oscillation, caused by opposing drives of local search and shape constraint.

In [CTL94], where ASMs were first introduced, it is suggested to detect convergence of the search by some threshold of change. If the change of landmark positions in an iteration drops below this threshold, the search is considered converged. However, due to noise and patterns in the target image that are not captured by the model, the landmarks will oscillate even if the shape and pose parameters are already within satisfactory ranges. To avoid the search to go on infinitely, a maximum number of loops is established. But when this number of loops is reached, there is no telling whether the search was successful in finding accurate shape and pose parameters or not.

Therefore we suggest to compensate the effects of noise by applying a gauss-filter to the magnitudes of change of the individual landmarks. Furthermore, rather than using the euclidean distance between the former and the updated landmark position to assess the magnitude of change (or the amplitude of oscillation), we introduce a measure that is supposed to be more capable of reflecting the agreement between texture matching and shape constraint. Also, this measure is proposed as criterion for segmentation success, assuming that if the landmarks are close to their true positions, the agreement will be high and otherwise low.

**Search procedure** Finding parameters  $\mathbf{p}$  in parameter space  $\mathbb{P}$  that fit a shape model to a target image can be viewed as the minimization of an error

function of local image content and shape deformation:

$$e(\mathbf{p}) = e_T(\mathbf{p}) \circ e_S(\mathbf{b}) \quad (2.45)$$

where  $\mathbf{b}$  are the shape parameters, and  $\mathbf{p}$  is the parameter vector  $(X_t, Y_t, s, \theta, \mathbf{b})$  determining translation, scale, orientation and shape. The texture error  $e_T$  is the overall profile appearance error at the landmark positions and  $e_S$  is the shape error.  $e_T$  can be calculated as an accumulation of the individual landmark texture errors  $e_{lm}$  which are given by the Mahalanobis distance of the sample gray value profile  $\gamma_{lm}$  to the mean texture profile  $\bar{\mathbf{g}}_{lm}$ , i.e.,

$$e_{lm} = (\gamma_{lm} - \bar{\mathbf{g}}_{lm})^T \mathbf{S}_{lm}^{-1} (\gamma_{lm} - \bar{\mathbf{g}}_{lm}), \quad (2.46)$$

where  $\mathbf{S}_{lm}$  is the covariance of the gray value profile samples in the training set. For instance, we can build *outline texture vectors* by concatenation of the texture vectors of all landmarks, and define  $e_T$  as the mahalanobis distance to the mean outline texture vector. Assuming the textures independent between the landmarks, the resulting definition of  $e_T$  would be

$$e_T = \sum e_{lm}(\mathbf{p})_i^2 \quad (2.47)$$

i.e. the squared sum of the errors of the individual landmarks.

The shape error  $e_S$  reflects the plausibility of the deformation given by the shape parameters and can be derived as the distance to the mean shape  $\mathbf{b}_{\text{mean}} = \vec{0}$ , using a metric which accounts for the correlations of the training set through the standard deviations of the elements  $b_i$  of  $\mathbf{b}$ , e.g.

$$e_S(\mathbf{b}) = \sum e_{S,i}(b_i) \quad (2.48)$$

$$e_{S,i}(b_i) = c \times |b_i|^n = n^{-1} \times (3\sigma_i)^{1-n} \times |b_i|^n \quad (2.49)$$

The slope of  $e_S$  is governed by  $n$ . The constant  $c$  is chosen such that

$$\frac{de_{S,i}(b_i)}{db_i} \begin{cases} < 1 & \text{if } |b_i| < 3\sigma_i \\ > 1 & \text{if } |b_i| > 3\sigma_i \end{cases} \quad (2.50)$$

See figure 2.15.

Intuitively, we would like to accept a parameter vector  $\mathbf{p}^*$  as plausible segmentation of an object if both  $e_T(\mathbf{p}^*)$  and  $e_S(\mathbf{b}^*)$  are low. However, due to noise and variations in illumination and texture it is hard to define a threshold for  $e_T$ . Either, the chosen threshold will often be too restrictive and reject correct delineations due to increased texture errors resulting from noisy target images, or, the threshold will fail to reliably classify all cases of failure or inaccurate results. Therefore, instead of defining a threshold for  $e_T$ , we try to derive a relative error measure by assuming that  $\mathbf{p}^*$  is plausible to segment an object correctly, only if *both*  $e_T(\mathbf{p}^*)$  and  $e_S(\mathbf{b}^*)$  have values *close to* the minimum errors within the search range. And we argue, that a parameter vector  $\mathbf{p}^*$ , which both satisfies

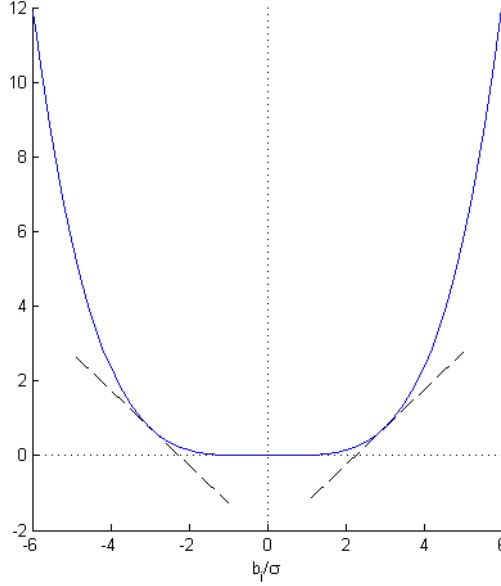


Figure 2.15: The shape error function  $e_S$ .

the shape constraint *and* yields a minimum texture error  $e_T(\mathbf{p}^*)$  within a sufficiently large search range is very likely to identify the target object in the image. This argument is encouraged by  $e_S$  and  $e_T$  being independent measures. Hence, letting  $d_T$  and  $d_S$  be some metrics (or at least quasisemimetrics), we assert

$$d_T(\mathbf{p}^*, \mathbf{p}_{min}) < t_T \quad (2.51)$$

and

$$d_S(\mathbf{b}^*, \mathbf{b}_{min}) < t_S \quad (2.52)$$

where  $\mathbf{b}_{min}$  and  $\mathbf{p}_{min}$  are the locations of the minimum shape error  $e_S$  and the minimum texture error  $e_T$ , respectively, within the search range. In the following we try to set up definitions of  $d_T$  and  $d_S$ .

Since  $e_S$  reflects the distance to the mean shape we know that the local and global minimum is found at  $\mathbf{b} = \vec{0}$  and has the value 0. We define

$$d_S(\mathbf{b}_1, \mathbf{b}_2) = |e_S(\mathbf{b}_1) - e_S(\mathbf{b}_2)| \quad (2.53)$$

so that

$$d_S(\mathbf{b}, \vec{0}) = e_S(\mathbf{b}). \quad (2.54)$$

As for shape deformation, we can thereby establish an absolute threshold  $t_S$  for  $e_S$ . Limiting the shape parameters  $\mathbf{b}$  will keep  $e_S(\mathbf{b})$  below this threshold and thereby maintain inequation 2.52, corresponding to a minimal acceptably posterior probability of the shape within the trained shape space, assuming a Gaussian distribution.

Note that there is no unique local minimum of  $e_T$ . Even within the search range, there can be more than one such minimum. The values of the local

minimum are unknown and  $e_T$  does not increase monotonically with increasing euclidean distance to the location  $\mathbf{p}_{\min}$  of the minimum error within the search range. Two points within the search range may lie far apart, but still fit the modeled texture similarly well, i.e.  $e_1(p_1) \approx e_1(p_2)$ . Vice versa, if the euclidean distance between two points in  $\mathbb{P}$  is small, they frequently offer very different fitness. As a first step, we can define

$$d'_T(\mathbf{p}_1, \mathbf{p}_2) = |e_T(\mathbf{p}_1) - e_T(\mathbf{p}_2)| \quad (2.55)$$

The advantage of this definition over the euclidean distance between the landmark positions is that it accounts for the *quality of fitness* rather than the offset from the optimal position. However, we still have to deal with absolute values of fitness, which may vary between images and even between regions within an image. To overcome this problem, the measure is normalized by dividing through the maximum difference of fitness between two points within the search range

$$d_T(\mathbf{p}_1, \mathbf{p}_2) = \left| \frac{e_T(\mathbf{p}_1) - e_T(\mathbf{p}_2)}{e_T(\mathbf{p}_{\max}) - e_T(\mathbf{p}_{\min})} \right| \quad (2.56)$$

As  $d_S(\mathbf{b}^*, \mathbf{b}_{\min})$  will be kept low by limiting the shape parameters, evaluating  $d_T(\mathbf{p}^*, \mathbf{p}_{\min})$  after shape constraint will yield an inverse measure of confidence in segmentation success. By replacing  $e_T$  by  $e_{lm}$ , one can analogously define a measure  $d_{lm}$  for each individual landmark, which allows to assign confidence values to regions along the contour (see Fig. 2.16).

The Fit Confidence  $F_{lm}(\mathbf{p})$  for an individual landmark can now be defined:

$$F_{lm}(\mathbf{p}) = 1 - d_{lm}(\mathbf{p}^*, \mathbf{p}_{\min}) \quad (2.57)$$

$$= 1 - \begin{cases} \frac{e_{lm}(\mathbf{p}) - e_{lm}(\mathbf{p}_{\min})}{e_{lm}(\mathbf{p}_{\max}) - e_{lm}(\mathbf{p}_{\min})} & \text{if } e_{lm}(\mathbf{p}_{\max}) \neq e_{lm}(\mathbf{p}_{\min}) \\ 0 & \text{otherwise} \end{cases} \quad (2.58)$$

where  $\mathbf{p}_{\max}$  and  $\mathbf{p}_{\min}$  are the locations of the maximum and minimum texture error  $e_{lm}$  within the search range. Although the overall Fit Confidence for the entire shape could be likewise derived from  $e_T$ , in this work we chose to derive it from the confidence values of the individual landmarks and calculated the overall Fit Confidence as

$$F(\mathbf{p}) = \sqrt{\sum \mathbf{F}_{lm}(\mathbf{p})^2 / \sqrt{n}}, \quad (2.59)$$

where  $n$  is the number of landmarks. In section 3.2 the capability of this measure to determine success of the search is experimentally tested.

### 2.5.3 Using Fit Confidence to control search

In the following 4 uses of fit confidence are proposed to improve the ASM search: 1. a better termination criterion for the search procedure, 2. individual search ranges for each landmark for the texture profile matching, 3. variable weights derived from fit confidence for the shape alignment, and 4. the possibility for re-initialization of the search procedure. All these improvements are based on the notion of fit confidence, and their effect will be evaluated in Sec. 3.

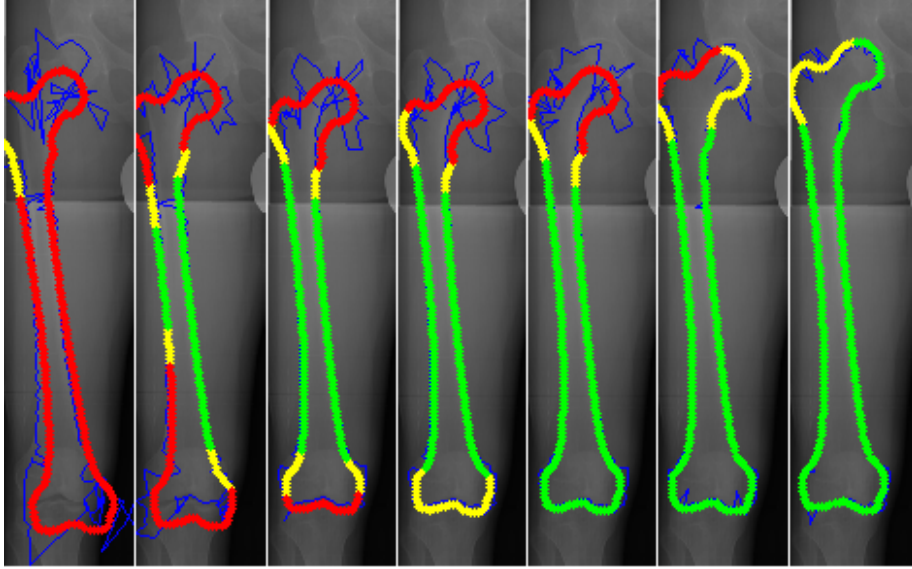


Figure 2.16: Fit Confidence calculated locally for individual landmarks. The images show successive iterations of the search procedure. The categories high, medium and low Fit Confidence are represented by green, yellow and red outline color, respectively.

### 2.5.3.1 Termination of the search procedure

In the original Active Shape Models formulation it was suggested to formulate the termination criterion for the iterative search algorithm in terms of a threshold of change. The procedure was to be repeated until no significant changes result [CTCG95]. However, over the iterations of the search procedure, the model points approach their correct positions in a non-monotonic manner, and the difference in the positions of the landmarks of two successive iterations does not decrease monotonically. In fact, even after iterations which caused only very little change, the approach can "pick up speed" and move the landmarks to their final positions. Furthermore, in some cases the model points are moved to their correct positions in only small steps. It is therefore not always possible to define an appropriate threshold of change to establish an optimal termination criterion. If the threshold is chosen too high, the search will terminate prematurely and provide inaccurate results. If it's chosen too low, the search will be unnecessarily prolonged, so that the time needed to locate the object will increase without improvement of accuracy. In the worst case, the search will engage in an oscillating behavior and never terminate, unless a maximum number of iterations is defined. Figures 2.17, 2.18 and 2.19 show exemplary values recorded during the search for a femoral bone in an example image. The correlation between the change in the landmark positions (Fig. 2.17) and the error (Fig. 2.18) is rather low (correlation coefficient  $c = 0.17$ ). The fit confidence on the other hand (Fig. 2.19) is highly correlated with the error ( $c = -0.98$ ) and thus allows to formulate a better termination criterion by setting a threshold based on this value. The plot indicates that the fit confidence is better suited to

serve as a termination criterion, than the position difference between successive iterations.

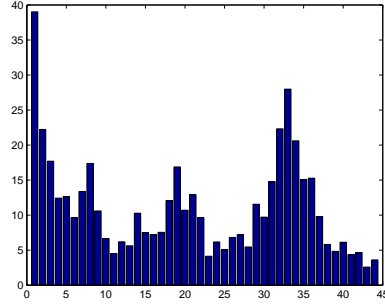


Figure 2.17: Euclidean distance between the landmark coordinate vectors of two successive iterations of the search procedure.

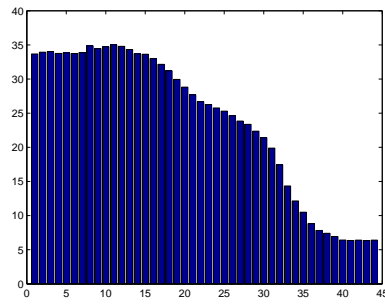


Figure 2.18: The maximum euclidean distance of a landmark to its correct position over the iterations of the search procedure.

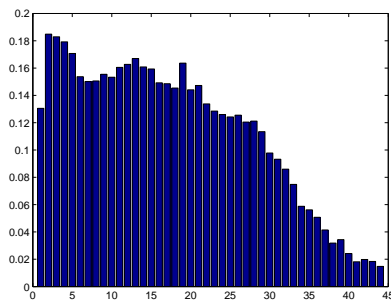


Figure 2.19: 1 - Fit Confidence over the iterations of the search procedure.

In the previous chapter, Fit Confidence has already been suggested as measure of confidence in success. Obviously, the problems mentioned above could be overcome, if the search algorithm terminated just by the time it succeeded in

attaining a desired accuracy. Thus, in the course of this work, Fit Confidence will be evaluated against change to detect that a desired level of accuracy has been reached. If Fit Confidence outperforms change in this regard, a novel termination criterion can be formulated to allow for more accurate segmentation results.

### 2.5.3.2 Variable search range

The search algorithm only works well when appropriate search ranges are used in the texture matching step (see the experimental results of increased and decreased ranges in Sect. 3.3). Due to the repetitive and cluttered structure of many anatomical sites a wide range can cause the divergence from the structure. Normalizing brightness and contrast and working on gradient images, as suggested in [CT04], even deteriorates this effect. In the case of overlapping structures, with multiple contours as e.g., in the hip region, this can make a reliable search for the structure of interest unfeasible. Even structures which apparently do not resemble the texture in question - e.g. seemingly solidly gray areas - come forward as top candidates of the texture matching algorithm. Occasionally the forces of the individual landmarks sum up to zero when the shape is updated, so that the landmarks engage in oscillation, i.e. they are infinitely moved to local texture error minima and immediately corrected back to their original locations when the landmark coordinates are projected onto the principal components of the model and the parameters are limited to preserve shape plausibility. In other cases, the shape will simply be moved out of the image. However, when the texture matching algorithm searches in too narrow a range, i.e. if the search range is smaller than the distance of a landmark to its correct location, the algorithm will not be able to move the landmark to the optimal position. Actually, the model point could even be moved farther away from it. The choice of the search range, especially in combination with the distance of the landmarks to their true positions, has therefore great influence on the accuracy of the segmentation result.

Starting the search with a wide search range and decreasing it over the iterations of the process will undoubtedly guarantee that the search will converge. However, the point in shape space where it will terminate is governed by the number of iterations needed to decrease the search range to zero, rather than by properties of the positions in the image to which the landmarks have been moved. Moreover, the correct landmark positions may simply move beyond reach when the search range is decreased monotonically.

For this application, an effective range control mechanism was developed and tested, which introduces individual dynamic search ranges for every model point, that are updated after every iteration, according to the current local fit confidence. High confidence will narrow the search range for the respective model point. Likewise, low confidence will widen the search range. If we define  $r_{min}$  and  $r_{max}$  to be the minimum and maximum search range, respectively, and  $F_{l,i}$  to be the local fit confidence of the landmark with index  $l$  after the  $i^{th}$  iteration, the search range  $r_{l,i+1}$  for this landmark in iteration  $i + 1$  is given

$$r_{l,i+1} = r_{max} - F_{l,i}(r_{min} - r_{max}) \quad (2.60)$$

Note that the value of  $F_{l,i}$  is in the range  $[0, 1]$  so that the search range will be

kept within the limits  $r_{min}$  and  $r_{max}$ .

### 2.5.3.3 Variable deformation limits and weights for shape alignment

Depending on the starting position of the search, on the visibility and image quality of the individual regions along the boundary of the object, the correct locations of some of the landmarks can more easily be found than those of others. For example, pathological variances in structure and appearance can make texture matching very difficult and fault-prone in the affected regions. In some cases the search algorithm has to cope with missing information, e.g. when parts of the bones are invisible, because the image is cropped or because they are occluded by screws, splints and plates in images which show post surgical status. The presence of a prosthesis (e.g. a femoral head prosthesis) causes the appearance of regions, which are crucial for the calculation of the mechanical axis, to deviate dramatically from the modeled texture. Figure 2.20 shows examples of such cases.

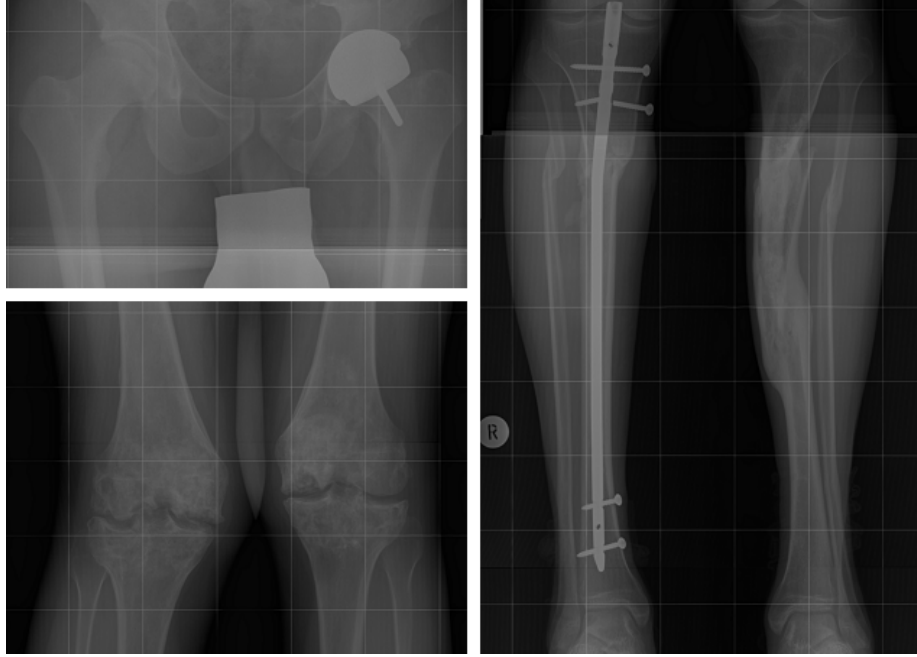


Figure 2.20: The radiographs are composites of 3 partial radiographs and show a quadratic measurement grid. Many images show post surgical status. Parts of the bones are occluded by screws, splints and plates. The presence of a prosthesis causes the appearance to deviate dramatically from the modeled texture.

In addition, due to the shape of the object, there are model points which cannot be matched to better positions until the landmarks are already close to their true locations. Figure 2.21 illustrates this.

The shape model allows the search to infer the positions of these landmarks from the remaining points. During search the weight of the contribution of individual points to the model parameter estimate can be adjusted in a con-



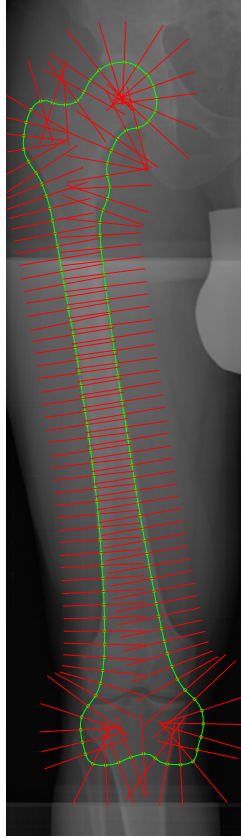


Figure 2.21: The local texture matching searches only perpendicular to the current contour (red lines) and, thus, fails to move the landmarks of the lower part of the bone to their correct positions.

tinuous manner, to account for their varying reliability. The search procedure can be roughly subdivided into two phases. During the first phase, the pose of the shape is still inaccurate, i.e. there are still considerable corrections of scale, translation and rotation necessary. The second phase is entered when the shape's pose is roughly aligned and the main corrections are with regard to the shape parameters.

Every iteration in the first phase, consisting of local texture matching and global shape constraint, can be viewed as a *vote* for the way the *pose* is updated. After every model point made it's contribution to translate, rotate or scale the shape, the prior knowledge about the object's shape constrains the result to a plausible shape with updated pose. This effect and the two phases of the search are inherent to and emerge from the standard ASM search procedure. However, it can be essentially reinforced by assigning higher weights to the contributions of landmarks which are already close to their correct locations during the first

phase and by explicitly entering the second phase, where the shape parameter constraint can be relaxed, when a required accuracy has been attained. Again, Fit Confidence can be used to estimate accuracy.

When updating the model parameters to fit the model to the new landmark positions, the first step is to find a translation  $dX_t, dY_t$ , a rotation  $d\theta$  and a scaling  $ds$  which best map the current set of points,  $\mathbf{X}$ , onto the set of updated landmark positions  $\mathbf{Y} = \mathbf{X} + d\mathbf{X}$ . This is done by finding the optimal parameters  $dX_t, dY_t, d\theta, ds$  to align  $\mathbf{X}$  and  $\mathbf{Y}$  so as to minimize the weighted sum-of-squares measure of point difference

$$E = \sum_{i=1}^n (\mathbf{Y} - T_{dX_t, dY_t, d\theta, ds}(\mathbf{X}))^\top \mathbf{W} (\mathbf{Y} - T_{dX_t, dY_t, d\theta, ds}(\mathbf{X})) \quad (2.61)$$

For our application, we choose  $\mathbf{W}$  to be a diagonal matrix with

$$w_{ll} = F_l \quad (2.62)$$

This way, landmarks with higher Fit Confidence are weighted higher in order that they drive the other landmarks to plausible positions.

#### 2.5.3.4 Re-initialization by interposition of an alternative search strategy

To compensate the by far most frequent effects of poor initialization, the contour of the bone is stretched and translated along its anatomical axis. This is done in a way that maximizes the cumulative Fit Confidence of all landmarks. Similarly to the fashion of the Life-wire algorithm explained in section 2.1.2, the cheapest path through a cost matrix  $\mathbf{C}$  is calculated. The row indices  $p$  of this matrix represent the indices of the model landmarks, whereas the column indices  $l$  reflect the lengths of individual translation vectors. The values of the elements of this matrix, i.e. the *costs*, are given by the Fit Confidence of the respective model point after translation:

$$C_{p,l} = 1 - F_l. \quad (2.63)$$

Fig. 2.22 depicts the cost map graphically. The cheapest path through all landmarks chooses translation coefficients for the individual model points. Translating the landmarks according to this path, guarantees that neighboring model points are translated similarly - their translation coefficients can differ by at most 1 pixel (except for the first and the last landmark, which can have a higher offset). Thereby, shape plausibility can be extensively maintained. Hence, the cheapest path through the cost map takes into account both shape deformation and overall texture fitness to translate the shape in an optimal way (see Figs. 2.22 and 2.23).

#### 2.5.4 Using sub-models and super-models

All the defining landmarks of the mechanical axes reside in regions around the joints, the correct segmentation of which is thus crucial for the accurate

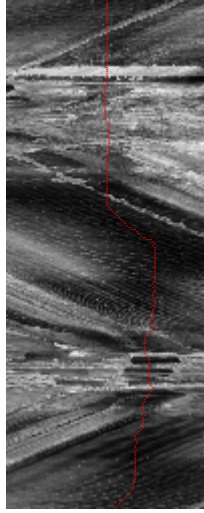


Figure 2.22: The cost map is made up by the negative local Fit Confidence of each landmark (rows), given they are translated by a certain amount (columns) in the direction of the anatomical axis of the bone.

identification of the axes. Therefore, *sub-models* are trained for the femoral head, the proximal and distal side of the knee joint and the talocrural joint. This way, the search will be less restricted by global shape model of the entire bone which allows the models to better fit to the actual contours in the target image. Also, the landmarks that form the contour will be denser which allows for more precise annotation of the mechanical axes.

While sub-models can refine delineation in regions of interest, a *super-model* can be used to model the global configuration of the limb bones in order to constrain their relative size, position and alignment w.r.t. each other. The global configuration is modeled by building a statistical shape model of all four bones, analogously to the way shape models were built for the individual bones as described in Sect. 2.2.1. 64 landmarks were used for each of the bones to represent its outlines. The global configuration constraint yields another measure of confidence. Implausible configurations like overlapping bones or disjointed femora and tibiae can be detected and the ASM search can be reinitialized with the “nearest plausible configuration”, which further improves the robustness of the results. When updating the pose parameters to fit the models to the “nearest plausible configuration”, the model points are weighted with the Fit Confidence of the respective bone according to the scheme described in Sect. 2.5.3.3 in order to estimate plausible positions of the bones with low confidence from those with higher confidence.

[RCA03] presented a similar approach of combining a global model with

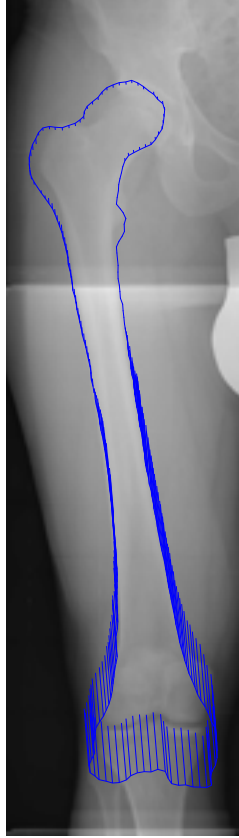


Figure 2.23: The "cheapest" path through all landmarks chooses translation coefficients for the individual model points, to move them to positions which maximize the cumulative Fit Confidence, while neighboring model points are always translated similarly.

a sequence of sub-models, where the global model applies iteratively updated "soft constraints" on the sub-models. They applied the algorithm to dual x-ray absorptiometry scans of the spine in order to automate vertebral morphometry measurements, using overlapping triplets of vertebrae as the sub-models, together with a global model of the entire spine, and reported substantially better results than using a single model.

## 2.6 Identifying the measuring points and calculating the axis angle

The landmarks defining the mechanical axes can be located in a straightforward way, when they are directly represented by model points. These are identified by their index and can be matched between different forms. In the course of the search, the correspondence of the model points is largely preserved, so that the landmark positions of the shape found in the image can be expected to render their true positions quite accurately. Hence, the landmarks that are to define the mechanical axes can be annotated manually in the training phase on any instance of the shapes. In this work, the landmarks were drawn from prototypical shape instances, which can be generated by setting the model parameters to their mean values (see figure 2.24). The three points  $h_1, h_2, h_3$  needed to calculate the center of the femoral head (Fig. 2.6) are selected manually, along with the center  $k_1$  of the intercondylar notch (Fig. 2.26), the center  $k_2$  of the tibial spines (Fig. 2.6) and the center  $a$  of the ankle mortise (Fig. 2.28).

After having located the points  $(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{k}_1, \mathbf{k}_2, \mathbf{a})$  the alignment angle is calculated. As a first step, the center  $\mathbf{h}_c$  of the femoral head is computed as the triangle circumcenter of  $(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)$ ,

$$\mathbf{h}_c = \mathbf{h}_1 + \frac{\begin{pmatrix} (h_{3y} - h_{1y})\mathbf{h}_1^\top \mathbf{h}_2 & - & (h_{2y} - h_{1y})\mathbf{h}_1^\top \mathbf{h}_3 \\ - & (h_{3x} - h_{1x})\mathbf{h}_1^\top \mathbf{h}_2 & + & (h_{2x} - h_{1x})\mathbf{h}_1^\top \mathbf{h}_3 \end{pmatrix}}{2((h_{3y} - h_{1y})(h_{2x} - h_{1x}) - (h_{2y} - h_{1y})(h_{3x} - h_{1x}))}. \quad (2.64)$$

The center  $k_c$  is calculated as the center between  $k_1$  and  $k_2$ ,

$$k_c = \frac{k_1 + k_2}{2}. \quad (2.65)$$

The mechanical axis angle is computed as

$$A = \pm(180 - \arctan^*(h_{cy} - k_{cy}, h_{cx} - k_{cx}) - \arctan^*(a_y - k_{cy}, a_x - k_{cx})), \quad (2.66)$$

where

$$\arctan^*(y, x) = \begin{cases} \arctan(\frac{y}{x}) & \text{if } x \geq y \\ \arctan(\frac{y}{x}) + \pi & \text{otherwise} \end{cases} \quad (2.67)$$

Note that the sign of the angle  $A$  depends on whether the left or right side is assessed and how varus and valgus alignment is to be encoded, respectively.

## 2.7 Summary of the image interpretation process

The algorithm for the interpretation of the image and assessment of the alignment angle can be summarized as follows:

1. Coarsely localize the four bones

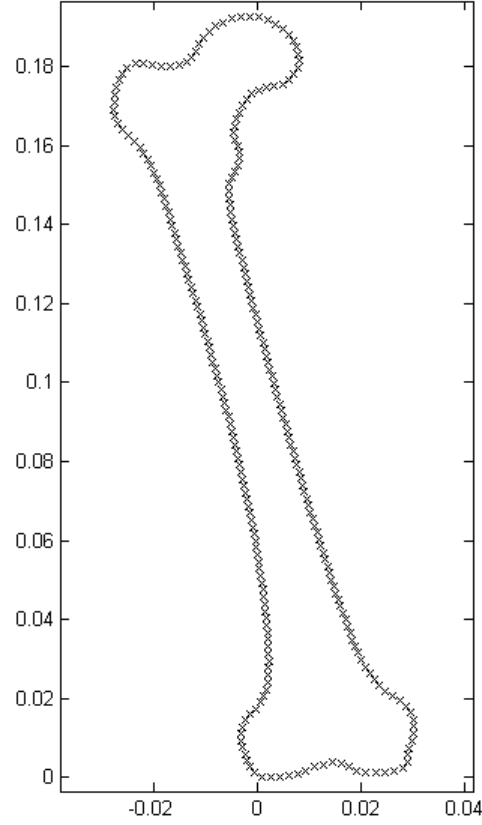


Figure 2.24: Prototypical instance of the right femur, generated by setting the model parameters to their mean values.

2.  $r \leftarrow 0$
3. For all four bones
  - (a) Perform improved multiresolution ASM search, controlled by estimations of Fit Confidence
  - (b) If Fit Confidence is too low
    - Reinitialize the search by stretching and translating the model along its anatomical axis so as to maximize Fit Confidence
    - Repeat step 3a
  - (c) Refine delineation by performing improved ASM search for both the proximal and distal joint region of the bone
4. Assess plausibility of the global configuration of the four bones.

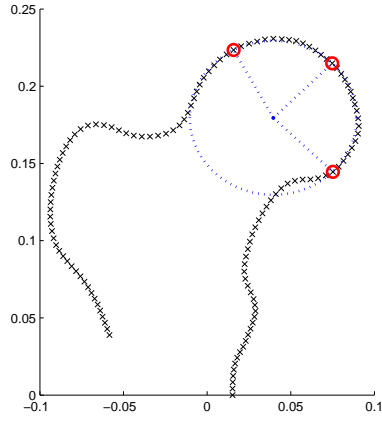


Figure 2.25: Prototypical instance of the right femoral head with the three points needed to calculate its center.

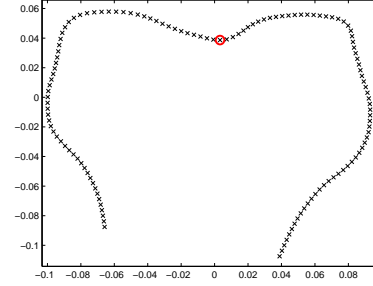


Figure 2.26: The center of the intercondular notch annotated on a prototypical instance of the proximal side of the right knee joint.

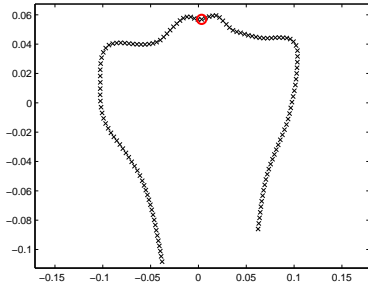


Figure 2.27: The center of the tibial spines annotated on a prototypical instance of the distal side of the right knee joint.

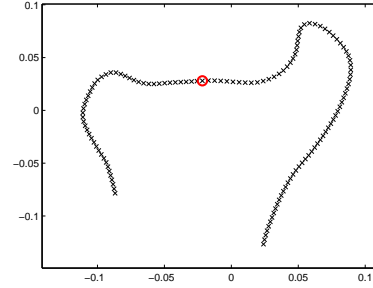


Figure 2.28: The center of the ankle mortise annotated on a prototypical instance of the right ankle joint.

5. If the global configuration is implausible and  $r < r_{max}$ 
  - $r \leftarrow r + 1$
  - Reestimate the initial positions of the bones by finding the “nearest plausible” configuration and update the model parameters accordingly, whereas the model points are weighted by the Fit Confidence of the respective bone in order to estimate the positions of bones with low confidence from bones with higher confidence
  - Continue with 3
6. Identify the landmarks defining the mechanical axes and calculate their angles

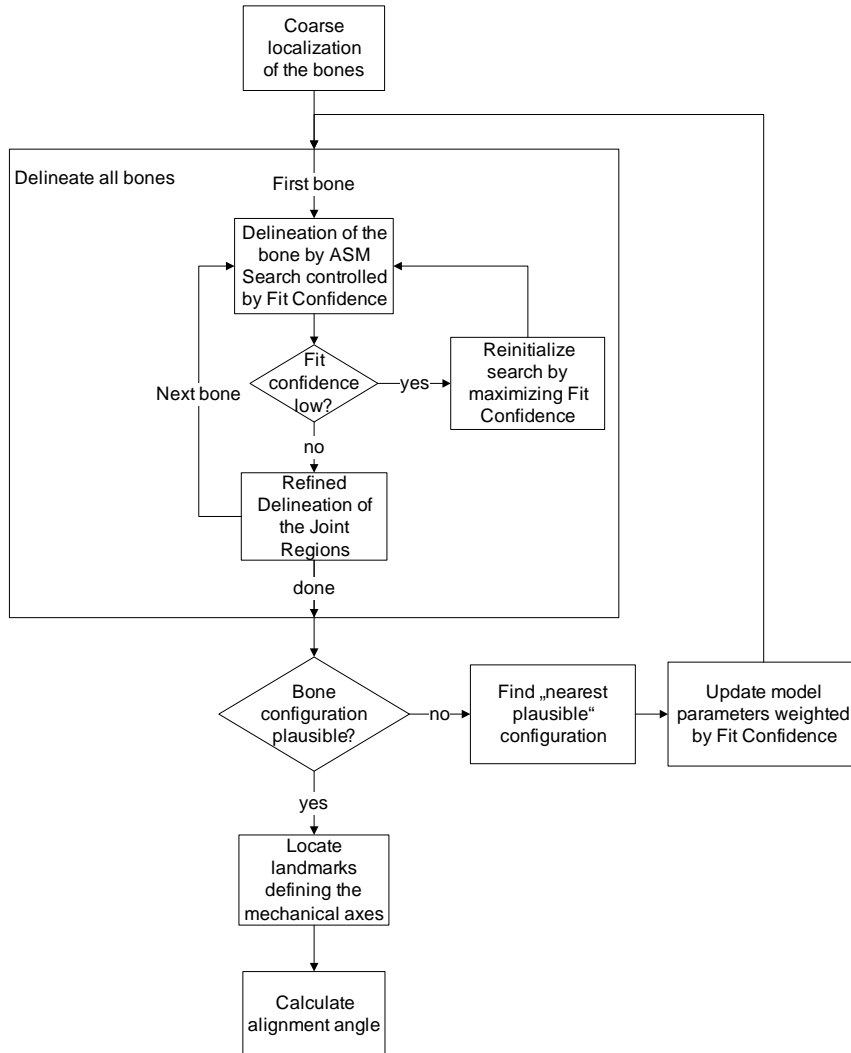


Figure 2.29: Schematic illustration of the image interpretation process

7. Return the angles along with the plausibility measure of the global configuration of the bones and the cumulative Fit Confidence of all bones and joints

A schematic overview of the image interpretation process is given in Fig. 2.29.



## Chapter 3

# Experiments

This Chapter reports on the experimental evaluation of the new methods. Three different aspects are tested separately: in Sect. 3.2 the power and specificity of failure detection based on *fit confidence* is assessed. Sect. 3.3 compares the standard ASM search and the enhanced search procedure introduced in Sect. 2.5.3 with regard to the accuracy of the results. Finally, Sect. 3.4 focuses on the functional evaluation of the automatic knee alignment measurements, including the agreement of the results with the standard of reference and measurement repeatability.

### 3.1 Common Setup

Experimental evaluation was performed on 26 full leg radiographs. The spatial resolution of the data was 0.3mm/pixel. The radiographs were composites of 3 partial radiographs and showed a quadratic measurement grid (see Fig. 2.20). For all images a manual expert annotation of the contours of the femora and the tibiae served as standard of reference. The training of the ASMs was performed at 15 radiographs and the testing on the remaining 11 radiographs (22 angles to measure and 44 bones to segment).

### 3.2 Fit Confidence

This Section reports on the evaluation of the power and specificity of failure detection based on *fit confidence*.

#### 3.2.1 Experimental setup

To test whether the Fit Confidence can be used to detect segmentation success and failure, respectively, the error rates of a success/failure detector based on Fit Confidence  $F(\mathbf{p})$  were computed from 1924 samples of parameter vectors  $\mathbf{p}$  of shape and pose. All of these parameter vectors were sampled during the ASM segmentation procedure after a random number of iterations. The Fit

Confidence was calculated as described in Sec. 2.5.2. The standard of reference regarding segmentation success was defined by a threshold  $t_d$  of the mean euclidean distance  $d$  of the landmark positions  $\mathbf{l}^* = (l_{x,1}^*, l_{y,1}^*, \dots, l_{x,n}^*, l_{y,n}^*)$  to their true locations  $\mathbf{l} = (l_{x,1}, l_{y,1}, \dots, l_{x,n}, l_{y,n})$ , i.e.

$$d(\mathbf{l}, \mathbf{l}^*) = \frac{1}{n} \sum_i \sqrt{(l_{x,i}^* - l_{x,i})^2 + (l_{y,i}^* - l_{y,i})^2} \quad (3.1)$$

where  $\mathbf{l}$  and  $\mathbf{l}^*$  are vectors of the x- and y-coordinates of the landmarks. The landmark errors  $d$  of the samples ranged from 0.64 to 41.15 pixels.

For our tests, we introduce a detector of *successful* segmentation and define the null hypothesis  $H_0$  to be that the delineation *has not* reached the desired accuracy. The rates of *false-negative* and *false-positive* results were evaluated for a threshold  $t_d = 10$  of distance and varying thresholds  $t_F$  of Fit Confidence.

### 3.2.2 Results

Figs. 3.1 and 3.2 show the distribution of  $d$  given  $F$  is above or below various values of threshold  $t_F$ .

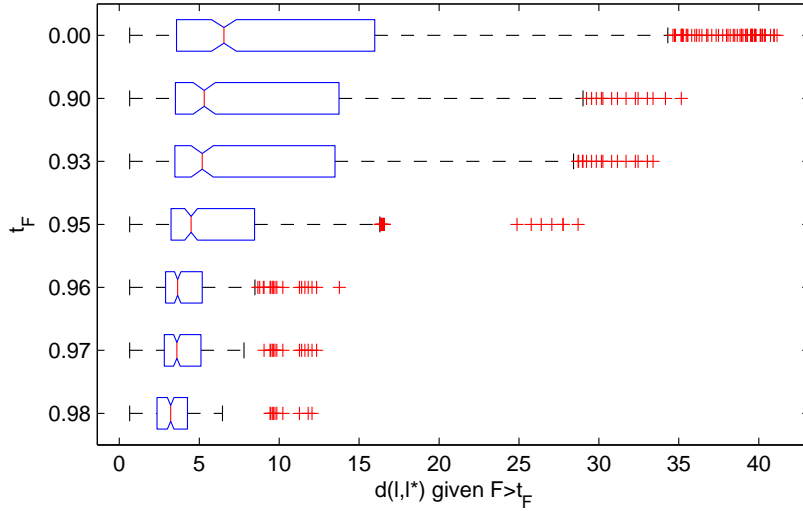


Figure 3.1: The distribution of the mean euclidean distance of the landmarks to their true positions given  $F > t_F$  with varied thresholds  $t_F$  of Fit Confidence.

Figure 3.3 shows the rates of *false-negative* and *false-positive* results for a threshold  $t_d = 10$  of landmark error and varying thresholds  $t_F$  of Fit Confidence.

Choosing  $t_F = 0.96$  resulted in 97% of the parameter vectors  $\mathbf{p}$  with  $F(\mathbf{p}) > t_F$  having landmark errors  $d < 10$  pixels (3% falsely accepted segmentation results). 84.2% of the remaining parameter vectors with  $F(\mathbf{p}) \leq t_F$  have landmark errors  $d > 10$  pixels (15.8% falsely rejected segmentation results). The distribution of the landmark errors of these parameter vectors can be seen in Fig. 3.1 and Fig. 3.2, respectively, at  $t_F = 0.96$ .

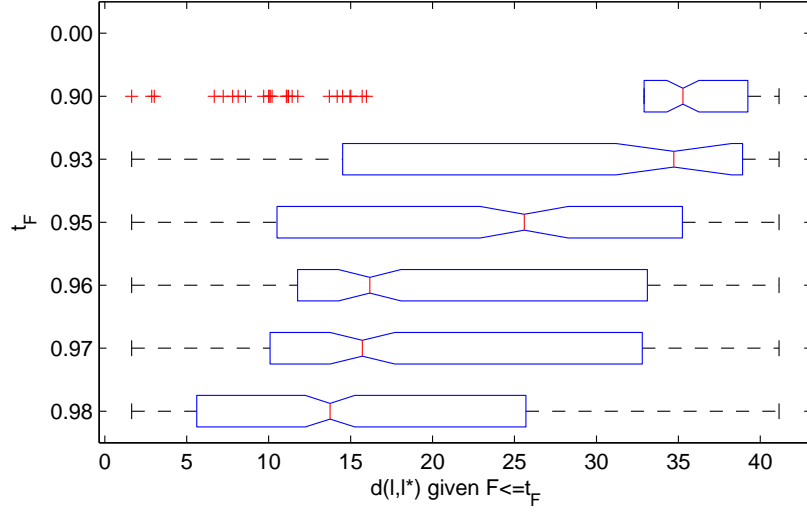


Figure 3.2: The distribution of the mean euclidean distance of the landmarks to their true positions given  $F \leq t_F$  with varied thresholds  $t_F$  of Fit Confidence.

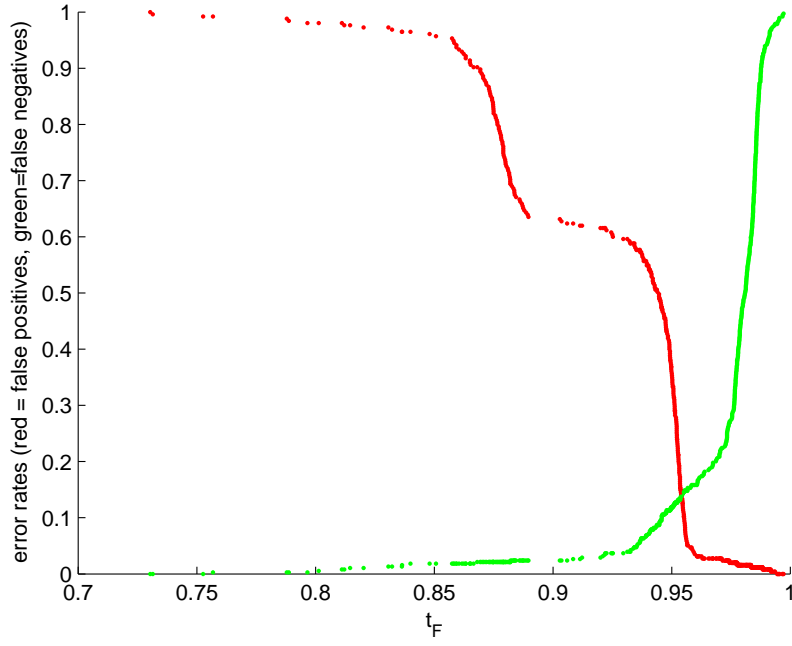


Figure 3.3: The error rates for a distance threshold  $t_d = 10$  and varying thresholds  $t_F$  of Fit Confidence. Every dot denotes one of 1924 samples.

### 3.2.3 Discussion

Calculating Fit Confidence has two purposes. Firstly, a mechanism is needed, which can reliably inform about possible failure. Secondly, high confidence in having reached a desired accuracy serves as a termination criterion for the search procedure. The experimental results show, that choosing the threshold of fit confidence optimally, only 3% segmentation results were falsely accepted and 15.8% of the results were falsely rejected, when a segmentation is defined to be successful if the landmark error  $d$  is below 10 pixels. This allows to reliably inform a human expert about low confidence in the results and a fairly high probability of failure. The low rate of falsely accepted results encourages the formulation of a termination criterion of the search procedure based on fit confidence.

Generally, the distribution of the landmark errors highly depends on the Fit Confidence. Hence, thresholding Fit Confidence yields an effective predictor of success and failure of the segmentation.

## 3.3 Search performance

In this Section, the enhancements of the search procedure proposed in Sect. 2.5.3 are evaluated. Their influence on the accuracy of the results is assessed by comparing the resulting distributions of the landmark errors.

### 3.3.1 Experimental setup

Active Shape Model search was performed with and without the use of Fit Confidence. To assess the error of the resulting delineation, the mean euclidean distance between the model landmarks and the standard of reference contour landmarks was evaluated. The proposed enhancements of the search procedure were evaluated by switching on and off the respective algorithm modifications and comparing the results.

In all, 7 runs were performed with varied settings:

1. the standard ASM search algorithm, with decreased search range (-10 pixels on either side)
2. the standard ASM search algorithm, with optimal search range
3. the standard ASM search algorithm, with increased search range (+10 pixels on either side)
4. ASM with variable search ranges
5. ASM with local fit confidences as weights for shape alignment
6. ASM with re-initialization triggered by fit confidence
7. ASM with all proposed adaptations

### 3.3.2 Results

Fig. 3.5 shows the distribution of the landmark errors resulting from the individual runs. With an optimal fixed search range, the basic ASM search procedure yields a mean error of  $7.86 \pm 14.22$  (min 0.07, max 72) pixels. The mean landmark error dropped to  $3.71 \pm 7.56$  pixels (min 0.01, max 40.34) when the search ranges were dictated by the local Fit Confidences of the individual landmarks. Using local Fit Confidences as weights for shape update, the mean landmark error was decreased to  $3.1 \pm 6.35$  pixels (min 0.03, max 46.57). Letting low Fit Confidence trigger re-initialization through a Fit Confidence maximizing procedure yielded a mean error of  $2.3 \pm 4.82$  pixels (min 0.01, max 40.34). With all proposed adaptations switched on, the mean error decreases to  $2.3 \pm 5.0$  pixels.

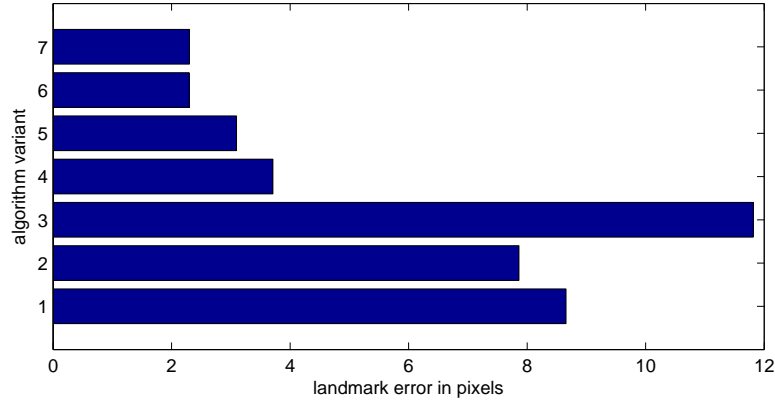


Figure 3.4: The mean landmark errors resulting from the different algorithm variants.

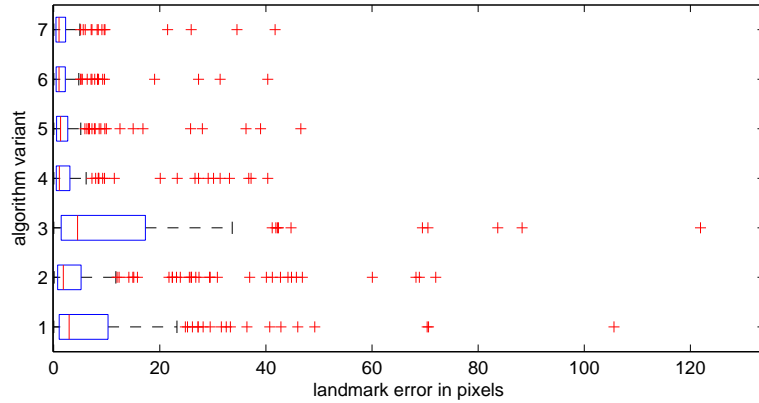


Figure 3.5: Distribution of the landmark errors resulting from the different algorithm variants.

### 3.3.3 Discussion

Using Fit Confidence to control the ASM search procedure improved the results significantly. The experimental results show, that each of the proposed modifications yields a better result accuracy as compared to the standard search procedure. Re-initialization of the search governed by fit confidence proved to have the strongest effect on the accuracy of the results. Actually, the combination of all proposed modifications could not further reduce the mean landmark error.

## 3.4 Angle measurement

In this Section the experiments and their results are presented, which assess the accuracy and repeatability of the automatic angle measurements.

### 3.4.1 Experimental setup

Accuracy is measured as the agreement between the results of the automatic assessment method and the results calculated from manually annotated landmarks, which serve as standard of reference. Calculating a meaningful measure of repeatability of a fully automatic method is inherently problematic. In the literature, repeatability is usually reported as the variance in the results of multiple measurements on the same image, either performed by the same or by different readers. However, provided exactly the same image, the results of repeated fully automatic measurements will always be identical, as the algorithm is perfectly deterministic. Yet, repeatability shall in practice not only be a measure of similarity between the results of multiple assessments on identical image data. Rather, we want the results to be stable against changes in the image data, that do not affect the quantity to be measured, i.e. the mechanical axis angle. More specific, there should always be a high agreement between the angles measured on two different radiographs of the same patient taken at the same time, since this is a prerequisite for the detection of small change. As the projection of the three-dimensional bones onto a 2D plane causes a loss of geometrical information, perfect agreement is theoretically impossible, unless the angle between the x-ray beams and the bone surfaces can be kept constant. Therefore, a second series of images was generated by clipping the original images varyingly. Thereby, a second image of every patient is obtained, which shows the lower limb bones from exactly the same perspective. To investigate the stability of the results against these modifications of the input image, the angles measured on the original images were compared to those measured on the second series. We argue, that we can thereby estimate a measure of the method’s reproducibility, that can be compared to the reported reproducibility of other methods.

For statistical evaluation, *leave-one-out cross validation* (LOOCV) was performed on the original training set. Further, an additional independent validation set was assembled, which consisted of radiographs which were not included in training.

The original training set was gathered from total-leg radiographs of 15 patients, 12 male and 3 female, with a mean age of  $29.4 \pm 6.9$  years. The alignment angles ranged from  $12.63^\circ$  (varus) to  $-10.14^\circ$  (valgus). For LOOCV, training was performed with data from 14 patients at a time, while the automatic segmentation was performed on the remaining case. Two of the right tibiae and one right femoral bone had to be excluded from training as they were either cropped or occluded to a large part by plates and screws, so that their outlines could not be reconstructed. Therefore, the two right knees affected were also excluded from LOOCV evaluation.

The independent validation set included full-limb radiographs of 11 patients (22 knees), 5 male and 6 female, with a mean age of  $28.2 \pm 5.9$  years. One of the patients showed mild, one showed grave deformations in at least one bone. Two of the radiographs showed a post surgical status with splints, screws or plates occluding parts of the bones. In one image parts of the bones were outside the image and hence cropped. The alignment angles of the test-set ranged from  $8.76^\circ$  (varus) to  $-9.49^\circ$  (valgus). For evaluation using the independent validation set, the complete original training set was used to build the models.

In order to test whether automatic measurement could replace manual assessment, we evaluated the difference between the results of the two methods. The alignment angles of the test-set were measured manually three times by a third-year radiology resident (six months of musculoskeletal subspecialty training). The measurements were conducted by locating the defining points of the mechanical axes on digital radiographs. The mean angles served as standard of reference for determining the accuracy of the automatic measurements. We calculated the mean, standard deviation, minimum and maximum of the differences and absolute differences between the measured angles. Furthermore, we assessed the *limits of agreement* and *bias* (with 95% CI), according to the methods suggested in [BA86], as well as the *minimal detectable change* as described in [GB07]. The differences are expected to be normally distributed, so that 95% of the differences will be found within these limits. To evaluate reproducibility, the automatic measurements were repeated with varying clipping (2% of image height and width, respectively) of the entire radiograph.

### 3.4.2 Results

Using the independent validation set, the mean absolute difference between the automatically measured angles from the standard of reference was  $0.53 \pm 0.44^\circ$  (min  $0.04^\circ$ , max  $1.57^\circ$ ). Fig. 3.6 shows the automatically assessed results plotted against the standard of reference. 82% of the angles differed less than  $1^\circ$  from the standard of reference, 59% differed less than  $0.5^\circ$ . The limits of agreement were  $-0.92^\circ$  ( $-1.32^\circ$  to  $-0.53^\circ$ , 95% CI) and  $1.52^\circ$  ( $1.12^\circ$  to  $1.92^\circ$ , 95% CI), with a bias of  $0.3^\circ$  ( $0.07^\circ$  to  $0.53^\circ$ , 95% CI) as depicted in Fig. 3.7. The repeated automatic measurements showed a mean difference (bias) of  $-0.1 \pm 0.52^\circ$  ( $-0.29^\circ$  to  $0.09^\circ$ , 95% CI) and limits of agreement of  $0.93^\circ$  ( $0.59^\circ$  to  $1.26^\circ$ , 95% CI) and  $-1.12^\circ$  ( $-1.46^\circ$  to  $-0.79^\circ$ , 95% CI) between the trials (Fig. 3.8). The minimal detectable change was  $1.03^\circ$ . The mean absolute difference between the two automatic measurements was  $0.37 \pm 0.38^\circ$  (min  $0.03^\circ$ , max  $1.79^\circ$ ).

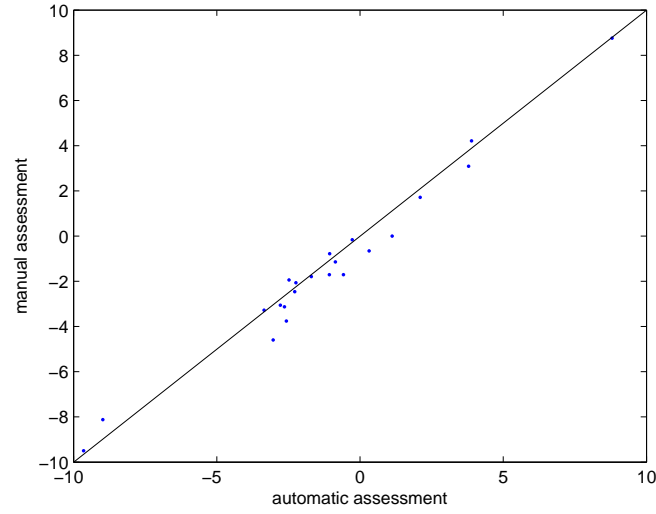


Figure 3.6: Results of the validation set: automatically assessed results plotted against standard of reference, with line of equality

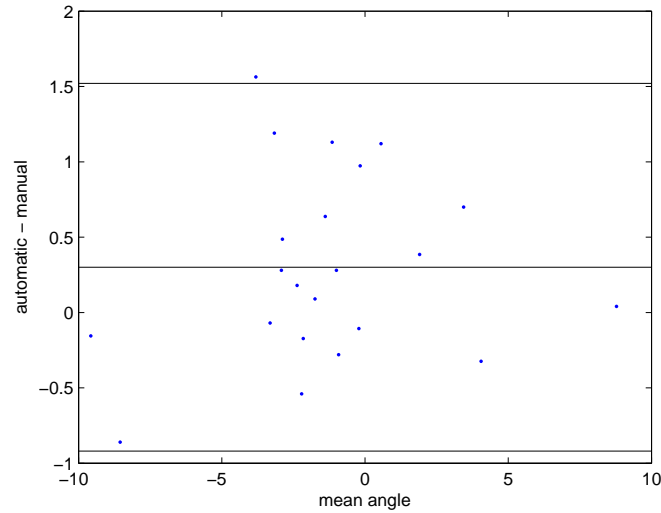


Figure 3.7: Results of the validation set: difference between automatic measurement and standard of reference, plotted against their mean, with bias and limits of agreement

Performing LOOCV, the mean absolute difference between the automatically measured angles from the standard of reference was  $0.58 + 0.46^\circ$  (min  $0.04^\circ$ , max  $1.69^\circ$ ). A plot of the automatically assessed results against the standard of reference can be seen in Fig. 3.9. 82% of the angles differed less than  $1^\circ$  from the standard of reference, 46% differed less than  $0.5^\circ$ . The limits of agreement were



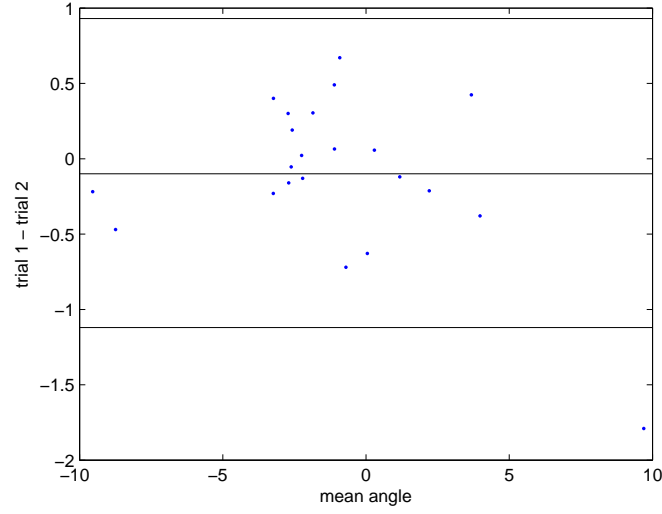


Figure 3.8: Results of the validation set: difference between automatic measurements with varied clipping of the radiograph against their mean. The black lines show the limits of agreement.

$-1.39^\circ$  ( $-1.81^\circ$  to  $-0.98^\circ$ , 95% CI) and  $1.53^\circ$  ( $1.11^\circ$  to  $1.94^\circ$ , 95% CI), with a bias of  $0.07 + 0.74^\circ$  ( $-0.17^\circ$  to  $0.31^\circ$ , 95% CI) as depicted in Fig. 3.10. The mean difference (bias) between repeated automatic measurements was  $0.04 + 0.19^\circ$  ( $-0.03^\circ$  to  $0.12^\circ$ , 95% CI) and limits of agreement of  $0.52^\circ$  ( $0.39^\circ$  to  $0.64^\circ$ , 95% CI) and  $-0.43^\circ$  ( $-0.56^\circ$  to  $-0.30^\circ$ , 95% CI) between the trials (Fig. 3.11). The minimum detectable change was  $0.47^\circ$ . The mean absolute difference between the two automatic measurements was  $0.16 + 0.19^\circ$  (min  $0.00^\circ$ , max  $0.81^\circ$ ).

### 3.4.3 Discussion

In [GB07] a reproducible, highly sensitive digital method to quantify knee alignment angle was reported. They evaluated longitudinal reproducibility of manual and digital measurements as well as their agreement concerning the measured angles. This was accomplished by the “limits of agreement” method described by [BA86]. The limits of agreement between two manual measurements performed on separate days were stated to be  $+1.65^\circ$  to  $-1.55^\circ$ , yielding a minimal detectable change of  $1.6^\circ$ . For the digital method, the limits of agreement were reported to be  $+0.43^\circ$  to  $-0.37^\circ$ , thus resulting in a minimal detectable change of  $0.4^\circ$ .

Deploying the method proposed in this thesis, the limits of agreement calculated from the validation set are  $0.93^\circ$  to  $-1.12^\circ$ , yielding a minimally detectable change of  $1.03^\circ$ . Performing LOOCV, the limits of agreement were  $0.52^\circ$  and  $-0.43^\circ$ , resulting in a minimal detectable change of  $0.47^\circ$ . Thus, this method outperforms the manual state of the art reported by [GB07] with respect to the agreement between repeated measurements. Since the 95% CI of the limits of agreement are  $0.39^\circ$  to  $0.64^\circ$  for the upper bound and  $-0.56^\circ$  to  $-0.30^\circ$  for the

	autom	msrmt	std of ref	diff	abs diff
left		0.32	-0.6531	0.98	0.98
		-0.27	-0.1638	-0.10	0.10
		-9.65	-9.4944	-0.15	0.15
		-1.70	-1.7902	0.09	0.09
		2.10	1.7151	0.38	0.38
		-2.64	-3.126	0.49	0.49
		-3.03	-4.5932	1.57	1.57
		-1.07	-1.7072	0.64	0.64
		3.79	3.0902	0.70	0.70
		-2.24	-2.0674	-0.17	0.17
right		3.89	4.2135	-0.32	0.32
		-0.86	-1.14	0.28	0.28
		-2.48	-1.94	-0.55	0.55
		-8.98	-8.12	-0.86	0.86
		8.80	8.76	0.04	0.04
		-1.06	-0.78	-0.27	0.27
		-0.58	-1.71	1.13	1.13
		-2.57	-3.76	1.19	1.19
		1.12	0.00	1.12	1.12
		-2.28	-2.46	0.18	0.18
	-2.78	-3.06	0.27	0.27	
	-3.35	-3.28	-0.08	0.08	
mean			0.30	0.53	
stddev			0.62	0.44	
min			-0.86	0.04	
max			1.57	1.57	

Table 3.1: Comparison of the automatically measured angles and the standard of reference, using an independent validation set.

	<b>trial 1</b>	<b>trial 2</b>	<b>diff</b>	<b>abs diff</b>
left	0.32	0.2626	0.06	0.06
	-0.27	0.359	-0.63	0.63
	-9.65	-9.4312	-0.22	0.22
	-1.70	-2.0044	0.31	0.31
	2.10	2.3132	-0.21	0.21
	-2.64	-2.5855	-0.05	0.05
	-3.03	-3.4308	0.41	0.41
	-1.07	-1.1344	0.07	0.07
	3.79	4.1691	-0.38	0.38
	-2.24	-2.2617	0.03	0.03
right	3.89	3.4661	0.43	0.43
	-0.86	-1.35	0.49	0.49
	-2.48	-2.67	0.19	0.19
	-8.98	-8.51	-0.47	0.47
	8.80	10.59	-1.79	1.79
	-1.06	-0.34	-0.72	0.72
	-0.58	-1.25	0.67	0.67
	-2.57	-2.87	0.31	0.31
	1.12	1.24	-0.12	0.12
	-2.28	-2.15	-0.13	0.13
	-2.78	-2.62	-0.16	0.16
	-3.35	-3.12	-0.23	0.23
<b>mean</b>			<b>-0.10</b>	<b>0.37</b>
<b>stddev</b>			<b>0.52</b>	<b>0.38</b>
<b>min</b>			<b>-1.79</b>	<b>0.03</b>
<b>max</b>			<b>0.67</b>	<b>1.79</b>

Table 3.2: Comparison of repeated automatic measurements, performed on an independent validation set.

	autom	msrmt	std of ref	diff	abs diff
left		-0.51	-0.21	-0.31	0.31
		0.96	0.90	0.07	0.07
		-0.24	-1.21	0.98	0.98
		-1.02	-1.06	0.04	0.04
		-7.13	-7.32	0.19	0.19
		-1.66	-1.61	-0.05	0.05
		-5.46	-4.74	-0.72	0.72
		-3.94	-2.68	-1.26	1.26
		3.52	3.48	0.04	0.04
		-2.71	-3.19	0.48	0.48
		3.72	2.74	0.98	0.98
		2.28	2.10	0.17	0.17
		7.95	7.11	0.84	0.84
		-2.30	-1.69	-0.61	0.61
right		0.40	0.93	-0.53	0.53
		-2.54	-2.80	0.26	0.26
		-3.44	-4.55	1.10	1.10
		-0.05	-0.76	0.71	0.71
		-9.30	-9.60	0.29	0.29
		0.67	0.63	0.04	0.04
		-10.07	-10.14	0.07	0.07
		-2.28	-2.14	-0.14	0.14
		12.04	12.63	-0.58	0.58
		-2.69	-3.53	0.84	0.84
		7.88	7.08	0.80	0.80
		0.86	-0.28	1.15	1.15
		-3.89	-2.64	-1.26	1.26
		-0.95	0.73	-1.69	1.69
mean				<b>0.07</b>	<b>0.58</b>
stddev				<b>0.74</b>	<b>0.46</b>
min				<b>-1.69</b>	<b>0.04</b>
max				<b>1.15</b>	<b>1.69</b>

Table 3.3: Comparison of the automatically measured angles and the standard of reference. The automatic measurements were conducted by performing a LOOCV.

	trial 1	trial 2	diff	abs diff
left	-0.51	-0.59	0.08	0.08
	0.96	0.86	0.10	0.10
	-0.24	-0.12	-0.12	0.12
	-1.02	-1.07	0.05	0.05
	-7.13	-7.38	0.25	0.25
	-1.66	-1.40	-0.27	0.27
	-5.46	-5.31	-0.15	0.15
	-3.94	-4.76	0.81	0.81
	3.52	3.54	-0.01	0.01
	-2.71	-2.83	0.11	0.11
	3.72	3.64	0.08	0.08
	2.28	2.08	0.20	0.20
	7.95	8.01	-0.06	0.06
	-2.30	-2.14	-0.17	0.17
	0.40	0.36	0.04	0.04
right	7.47	7.45	0.02	0.02
	-2.54	-2.59	0.05	0.05
	-3.44	-3.19	-0.25	0.25
	-0.05	-0.08	0.03	0.03
	-9.30	-9.26	-0.04	0.04
	0.67	0.68	-0.02	0.02
	-10.07	-10.26	0.19	0.19
	-2.28	-2.44	0.16	0.16
	12.04	11.96	0.09	0.09
	-2.69	-2.74	0.06	0.06
	-0.10	0.12	-0.22	0.22
	7.88	7.88	0.00	0.00
	0.86	0.14	0.72	0.72
	-3.89	-3.75	-0.14	0.14
	-0.95	-0.68	-0.27	0.27
<b>mean</b>			<b>0.04</b>	<b>0.16</b>
<b>stddev</b>			<b>0.24</b>	<b>0.19</b>
<b>min</b>			<b>-0.27</b>	<b>0.00</b>
<b>max</b>			<b>0.81</b>	<b>0.81</b>

Table 3.4: Comparison of repeated automatic measurements, conducted by performing a LOOCV.

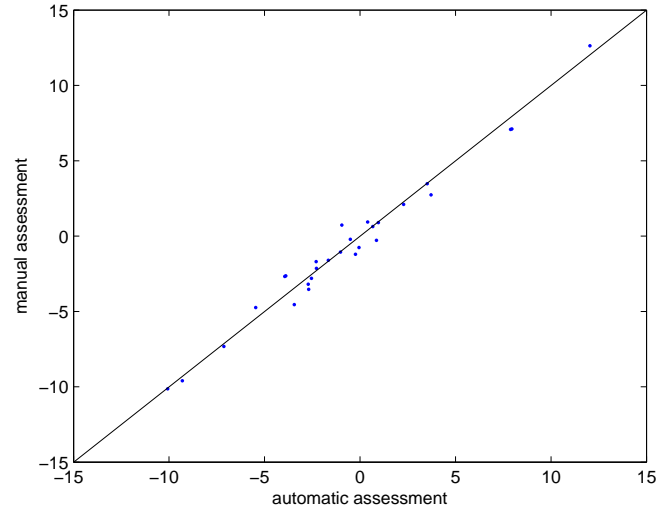


Figure 3.9: Results of the LOOCV: automatically assessed results plotted against standard of reference, with line of equality

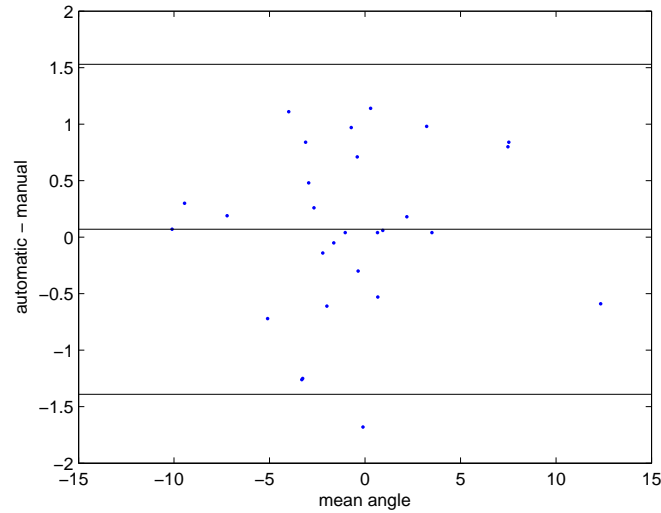


Figure 3.10: Results of the LOOCV: difference between automatic measurement and standard of reference, plotted against their mean, with bias and limits of agreement

lower bound, we can consider this to be comparable to the results of the digital method presented by [GB07]. However, while this digital method involves a number of user interactions at different levels of the application, the results described in this thesis are achieved fully automatic.

The limits of agreement between the automatically measured angles and the

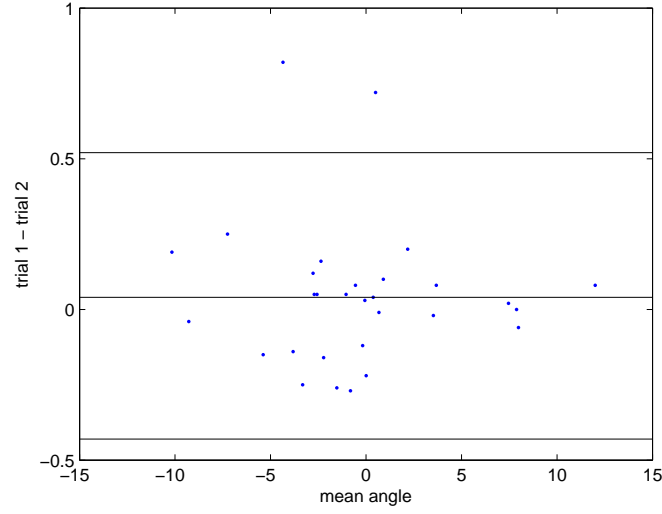


Figure 3.11: Results of the LOOCV: difference between automatic measurements with varied clipping of the radiograph against their mean. The black lines show the limits of agreement.

standard of reference are below the inter-reader variability reported in [GB07]. So, the automatically measured angles can replace manual readings without loss of accuracy.

The algorithm can cope with a certain amount of ambiguous image content, overlapping structures and missing data. However, the results generally depend on the “quality” of the image, where “bad quality” refers to the presence of a prosthesis, screws or plates, cropping and grave pathological deformations of the bones. The outlier in Fig. 3.7 (differences between repeated measurements) is caused by a deformation of the right tibia and the right ankle joint. This makes the algorithm unstable and the result highly depends on optimal initialization. The image where the automatically measured angle most deviates from the standard of reference in Fig. 3.9 shows a plate and screws in the left knee joint. The outliers in Fig. 3.10 are caused by a femoral head prosthesis and a large intramedullary pin, respectively.

## Chapter 4

# Conclusion

This thesis proposes a method to automatically measure the knee alignment angles on full-limb radiographs. Measuring the angles without any user interaction aims at improving accuracy and repeatability and thereby allow for detection of even small changes. High accuracy of the measured angles is a prerequisite for precise preoperative planning, for assessing the success of surgical correction and for the prediction of the clinical outcome.

The method involves three main steps. First, the positions of the bones in the target image are coarsely estimated using *Sparse Markov Random Field Appearance Models*. This technique integrates local descriptor similarities and deformation constraints in a single optimization step. It operates on configurations of symmetry-based interest points which are represented by graphs and Markov random fields, and performs matching by the MAX-SUM algorithm. As an alternative to the Sparse MRF Appearance Models, an application specific approach was used. In this approach the legs and the hip are segmented from the background using a filter. Calculating the medial axis transform of the result yields a “skeleton”, which approximates the location of the bones.

The next step is to use the resulting estimations to initialize a multi-resolution *Active Shape Model* (ASM) search. ASMs allow to utilize a priori knowledge about shape and appearance of the objects of interest. The search procedure is enhanced by strategies for increasing robustness by estimating fit confidence and asserting plausibility of the global configuration. These estimates are used to control the search and, if necessary, to trigger re-initialization procedures and repeat the search. The use of sub models allows to refine delineation of the joint regions, where all landmarks reside, that define the mechanical axes.

The final step is to identify these landmarks and calculate the alignment angles. The final estimates of fit confidence and plausibility of the global configuration of the bones are used for failure detection, which allows to inform the user about low confidence in the results.

To improve the result robustness, this thesis introduces enhancements of the standard ASM search procedure, which use the Fit Confidence to control the search. Experimental results show, that these enhancements improved the accuracy of the results significantly. Further, the distribution of the landmark



errors highly depends on the Fit Confidence. Hence, thresholding Fit Confidence yields an effective predictor of success and failure of the segmentation.

The method proposed in this thesis outperforms the manual state of the art reported by [GB07] with respect to the agreement between repeated measurements. The variability can be considered comparable to the results of the digital method presented by [GB07]. However, while this digital method involves a number of user interactions at different levels of the application, the results described in this thesis are achieved fully automatic.

The limits of agreement between the automatically measured angles and manual readings are below the inter-reader variability reported in [GB07]. Thus, the automatic angle measurements performed with the method introduced in this thesis can replace manual readings without loss of accuracy.

# Bibliography

- [BA86] J.M. Bland and D.G. Altman. Statistical method for assessing agreement between two methods of clinical measurement. *Lancet*, pages 307–310, 1986.
- [BA00] G.A. Brown and A. Amendola. Radiographic evaluation and pre-operative planning for high tibial osteotomies. *Oper Tech Sports Med*, 8:2–14, 2000.
- [BBL05] R. Beichel, H. Bischof, F. Leberl, and M. Sonka. Robust active appearance models and their application to medical image analysis. *IEEE TMI*, 24(9):1151–1169, 2005.
- [BM97] W.A. Barrett and E.N. Mortensen. Interactive live-wire boundary extraction. *Medical Image Analysis*, 1(4):331–341, 1997.
- [BMP01] Serge Belongie, Jitendra Malik, and Jan Puzicha. Matching shapes. In *ICCV*, pages 454–463, 2001.
- [Boo96] F. L. Bookstein. Landmark methods for forms without landmarks: Localizing group differences in outline shape. In *MMBIA '96: Proceedings of the 1996 Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA '96)*, page 279, Washington, DC, USA, 1996. IEEE Computer Society.
- [CET98] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *Lecture Notes in Computer Science*, 1407:484ff, 1998.
- [CHTH94] 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.
- [CIW93] MB Coventry, DM Ilstrup, and SL Wallrichs. Proximal tibial osteotomy. *J Bone Joint Surg*, 75:196–201, 1993.
- [CLS94] T.D. Cooke, J. Li, and R.A. Scudamore. Radiographic assessment of bony contributions to knee deformity. *Orthop Clin North Am.*, 25:387–393, 1994.
- [Coo04] T. F. Cootes. Timeline of developments in algorithms for finding correspondences across sets of shapes and images, June 2004.

- [CSB<sup>+</sup>91] TDV Cooke, RA Scudamore, JT Bryant, C Sorbie, D Siu, and B Fisher. A quantitative approach to radiography of the lower limb: principles and applications. *Journal of Bone and Joint Surgery*, 73B(5):715–720, 1991.
- [CSL<sup>+</sup>97] D Cooke, A Scudamore, J Li, U Wyss, T Bryant, and P Costigan. Axial lower-limb alignment: comparison of knee geometry in normal volunteers and osteoarthritis patients. *Osteoarthritis and Cartilage*, 5:39–47, 1997.
- [CT92] T. Cootes and C. Taylor. Active shape models- - smart snakes, 1992.
- [CT01] T.F. Cootes and C.J. Taylor. Statistical models of appearance for medical image analysis and computer vision. In *Proc. SPIE Medical Imaging*, 2001.
- [CT04] T.F. Cootes and C. Taylor. Statistical models of appearance for computer vision, 2004.
- [CTCG92] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Training models of shape from sets of examples, 1992.
- [CTCG95] T.F. Cootes, C.J. Taylor, D. Cooper, and J. Graham. Active shape models—their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.
- [CTL94] T.F. Cootes, C.J. Taylor, and A. Lanitis. Multi-resolution search with active shape models. *Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, 1:610–612 vol.1, 1994.
- [CWHW04] F Cicuttini, A Wluka, J Hankin, and Y Wang. Longitudinal study of the relationship between the knee angle and tibiofemoral cartilage. *Rheumatology*, 43:321–324, 2004.
- [DCT01] Rhodri H. Davies, Timothy F. Cootes, and Christopher J. Taylor. A minimum description length approach to statistical shape modelling. In *IPMI '01: Proceedings of the 17th International Conference on Information Processing in Medical Imaging*, pages 50–63, London, UK, 2001. Springer-Verlag.
- [DM98] I.L. Dryden and K.V. Mardia. *Statistical Shape Analysis*. John Wiley & Sons, 1998.
- [DML<sup>+</sup>07] Rene Donner, Branislav Micusk, Georg Langs, Lech Szumilas, Philipp Peloschek, Klaus Friedrich, and Horst Bischof. Object localization based on markov random fields and symmetry interest points. In Nicholas Ayache, Sbastien Ourselin, and Anthony J. Maeder, editors, *MICCAI (2)*, volume 4792 of *Lecture Notes in Computer Science*, pages 460–468. Springer, 2007.
- [DTC<sup>+</sup>02] 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. *MedImg*, 21(5):525–537, May 2002.

- [GB07] B. Goker and J. Block. Improved precision in quantifying knee alignment angle. *Clin Orthop Relat Res*, 2007.
- [Gow75] J. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, March 1975. available at <http://ideas.repec.org/a/spr/psycho/v40y1975i1p33-51.html>.
- [Gre95] R.P. Grelsamer. Unicompartmental osteoarthritis of the knee. *J Bone Joint Surg Am.*, 77:278–292, 1995.
- [HCT92] A. Hill, T.F. Cootes, and C.J. Taylor. A generic system for image interpretation using flexible templates. In *in Proc. British Machine Vision Conference*, pages 276–285. Springer-Verlag, 1992.
- [HD04] P. Hernigou and G. Deschamps. Alignment influences wear in the knee after medial unicompartmental arthroplasty. *Clin Orthop Relat Res*, 423:161–165, 2004.
- [HFB<sup>+</sup>06] M.A. HUNT, P.J. FOWLER, T.B. BIRMINGHAM, T.R. JENKYN, and J.R. GIFFIN. Foot rotational effects on radiographic measures of lower limb alignment. *Canadian journal of surgery*, 49(6):401–406, 2006.
- [HTB00] Andrew Hill, Chris J. Taylor, and Alan D. Brett. A framework for automatic landmark identification using a new method of non-rigid correspondence. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(3):241–251, 2000.
- [IDC<sup>+</sup>07] S.N. Issa, D. Dunlop, A. Chang, J. Song, P.V. Prasad, A. Guermazi, C. Peterfy, S. Cahue, M. Marshall, D. Kapoor, K. Hayes, and L. Sharma. Full-limb and knee radiography assessments of varus-valgus alignment and their relationship to osteoarthritis disease features by magnetic resonance imaging. *Arthritis Rheum*, 57(3):398–406, 2007.
- [KG92] C. Kambhamettu and D.B. Goldgof. Point correspondence recovery in non-rigid motion. *CVPR*, 92:222–227, 1992.
- [KNWT02] T. F. Cootes K. N. Walker and C. J. Taylor. Automatically building appearance models from image sequences using salient features. *Image and Vision Computing*, 20:435–440, apr 2002.
- [Kol06] Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10):1568–1583, 2006.
- [KPG90] KA Krackow, CL Pepe, and EJ Galloway. A mathematical analysis of the effect of flexion and rotation an apparent varus/valgus alignment an the knee. *Orthopedics*, 13:861–868, 1990.
- [KT98] Aaron C. Kotcheff and Chris J. Taylor. Automatic construction of eigenshape models by direct optimization. *Medical Image Analysis*, 2(4):303–314, December 1998.

- [KVWM05] V. B. Kraus, T. P. Vail, T. Worrell, and G. McDaniel. A comparative assessment of alignment angle of the knee by radiographic and physical examination methods. *Arthritis & Rheumatism*, 52(6):1730–1735, 2005.
- [KWT88] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, V1(4):321–331, January 1988.
- [LE77] PA Lotke and ML Ecker. Influence of positioning of prosthesis in total knee replacement. *The Journal of Bone and Joint Surgery*, 59A:77–79, 1977.
- [LPBK07] Georg Langs, Philipp Peloschek, Horst Bischof, and Franz Kainberger. Model-based erosion spotting and visualization in rheumatoid arthritis. *Acad Radiol*, 14(10):1179–1188, 2007.
- [Mah30] P.C. Mahalanobis. On tests and measures of groups divergence i. *Journal of the Asiatic Society of Benagal*, 26:541, 1930.
- [MBH87] JR Moreland, LW Bassett, and GJ Hanker. Radiographic analysis of the axial alignment of the lower extremity. *The Journal of Bone and Joint Surgery*, 69A(5):745–749, June 1987.
- [MMBU92] E. Mortensen, B. Morse, W. Barrett, and J. Udupa. Adaptive boundary detection using ‘live-wire’ two-dimensional dynamic programming. *Computers in Cardiology 1992. Proceedings.*, pages 635–638, 1992.
- [MT96] T. McInerney and D. Terzopoulos. Deformable models in medical images analysis: a survey, 1996.
- [Mur94] S.B. Murphy. Tibial osteotomy for genu varum: indications, pre-operative planning, and technique. *Othop Clin North Am.*, 25:477–482, 1994.
- [PE88] TL Petersen and GA Engh. Radiographic assessment of knee alignment. *J Arthroplasty*, 3:67–72, 1988.
- [PLW<sup>+</sup>07] P Peloschek, G Langs, M Weber, J Sailer, M Reisegger, H Imhof, H Bischof, and F Kainberger. An automatic model-based system for joint space measurements on hand radiographs: Initial experience. *Radiology*, 245(3):855–862, 2007.
- [PT92] D Paley and K Tetsworth. Mechanical axis deviation of the lower limbs: preoperative planning of uniapicular deformities of the tibia or femur. *Clin Orthop*, 280:48–64, 1992.
- [RCA03] M. Roberts, T. Cootes, and J. Adams. Linking sequences of active appearance sub-models via constraints: an application in automated vertebral morphometry. in *Proc.BMVC*, 1:349–358, 2003.

- [RCB97] Anand Rangarajan, Haili Chui, and Fred L. Bookstein. The softassign procrustes matching algorithm. In *IPMI '97: Proceedings of the 15th International Conference on Information Processing in Medical Imaging*, pages 29–42, London, UK, 1997. Springer-Verlag.
- [Ris83] J. Rissanen. A universal prior for integers and estimation by minimum description length. *Ann. Stat.*, 11:416–431, 1983.
- [RMP<sup>+</sup>96] Anand Rangarajan, Eric Mjolsness, Suguna Pappu, Lila Davachi, Patricia S. Goldman-Rakic, and James S. Duncan. A robust point matching algorithm for autoradiograph alignment. In *VBC '96: Proceedings of the 4th International Conference on Visualization in Biomedical Computing*, pages 277–286, London, UK, 1996. Springer-Verlag.
- [SCB<sup>+</sup>91] D Siu, DV Cooke, LD Broekhoven, M Lam, B Fisher, G Saunders, and T Challis. A standardized technique for lower limb radiography: Practice, applications, and error analysis. *Investigative Radiology*, 26:71–77, 1991.
- [SLH91] G.L. Scott and H.C. Longuet Higgins. An algorithm for associating the features of two images. *RoyalP*, B-244:21–26, 1991.
- [SLT04] Charles V. Stewart, Ying-Lin Lee, and Chia-Ling Tsai. An uncertainty-driven hybrid of intensity-based and feature-based registration with application to retinal and lung ct images. In Christian Barillot, David R. Haynor, and Pierre Hellier, editors, *MICCAI (1)*, volume 3216 of *Lecture Notes in Computer Science*, pages 870–877. Springer, 2004.
- [SP95] Stan Sclaroff and Alex P. Pentland. Modal matching for correspondence and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(6):545–561, 1995.
- [SSF<sup>+</sup>01] L Sharma, J Song, DT Felson, S Cahue, E Shamiyeh, and DD Dunlop. The role of knee alignment in disease progression and functional decline in knee osteoarthritis. *JAMA*, 286:188–195, 2001.
- [SSP<sup>+</sup>04] J. Sailer, M. Scharitzer, P. Peloschek, A. Giurea, H. Imhof, and S. Grampp. Quantification of axial alignment of the lower extremity on conventional and digital total leg radiographs. *European Radiology*, 15(1):170–173, Jan 2004.
- [Ste00] M.B. Stegmann. Active appearance models: Theory, extensions and cases. Master’s thesis, Technical University of Denmark, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, 2000.
- [TR01] H.H. Thodberg and A Rosholm. Application of the active shape model in a commercial medicals device for bone densitometry. In T.F. Cootes and C.J. Taylor, editors, *Proceedings of British Machine Vision Conference*, pages 43–52, 2001.

- [USB92] J.K. Udupa, S. Samarasekera, and W. A. Barrett. Boundary detection via dynamic programming. In R. A. Robb, editor, *Proc. SPIE, Visualization in Biomedical Computing*, volume 1808 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 33–39, September 1992.
- [Wer05] Tomáš Werner. A linear programming approach to Max-sum problem: A review. Research Report CTU–CMP–2005–25, Czech Technical University, 2005.
- [WLD<sup>+</sup>08] P. Widhalm, G. Langs, R. Donner, N. Fakhrai, P. Peloschek, and R. Sablatnig. Estimation of fit confidence in active shape model search for the reliable measurement of knee alignment. In A. Kuijper, B. Heise, and L. Muresan, editors, *Challenges in the Biosciences: Image Analysis and Pattern Recognition Aspects*, pages 79–90. Austrian Association for Pattern Recognition (AAPR/OAGM), May 2008.
- [WWL06] Shuchang Wang, Yangsheng Wang, and Bai Li. Face decorating system based on improved active shape models. In *ACE '06: Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*, page 65, New York, NY, USA, 2006. ACM.
- [XP98] Chenyang Xu and Jerry L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Trans. on Image Proc.*, 7(3), March 1998.
- [YCTN07] W.P. Yau, K.Y. Chiu, W.M. Tang, and T.P. Ng. Coronal bowing of the femur and tibia in chinese: its incidence and effects on total knee arthroplasty planning. *Journal of Orthopaedic Surgery*, 15(1):32–36, 2007.
- [YLL<sup>+</sup>03] S. Yan, C. Liu, S. Li, H. Zhang, H. Shum, and Q. Cheng. Face alignment using texture-constrained active shape models, 2003.