Anatomical Shape and Motion Reconstruction from Sparse Image Data
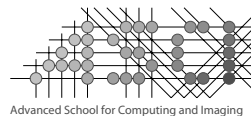
Nóra Baka

NWO

Netherlands Organisation for Scientific Research

Advanced School for Computing and Imaging

Anatomical Shape and Motion Reconstruction from Sparse Image Data

Anatomische vorm- en bewegingsreconstructie op basis van schaarse beelddata

---

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR AAN DE

ERASMUS UNIVERSITEIT ROTTERDAM

OP GEZAG VAN DE RECTOR MAGNIFICUS

PROF.DR. H.G. SCHMIDT

EN VOLGENS BESLUIT VAN HET COLLEGE VOOR PROMOTIES.

DE OPENBARE VERDEDIGING ZAL PLAATSVINDEN OP
VRIJDAG 16 NOVEMBER 2012 OM 13.30 UUR DOOR

Nóra Baka

GEBOREN TE KAPOSVÁR, HONGARIJE

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotoren: Prof.dr. W.J. Niessen
Prof.dr.ir. B.P.F. Lelieveldt

Overige leden: Prof.dr. H. Weinans
Prof.dr. B. van Ginneken
Dr. G. Penney

Copromotor: Dr. M. de Bruijne

Contents

Chapter 1

Introduction

1.1 Sparse imaging in current medical practice

In current clinical practice, medical imaging plays a key role in diagnosis, therapy planning and therapy monitoring. Some of these modalities, such as CT, MRI, and 3D ultrasound, provide high resolution volumetric anatomical information, and more recently, 3D imaging in time. In certain practical situations, however, limitations with respect to imaging time, space, radiation dose, or ergonomics make it impossible to acquire such rich data. In such cases, imaging may be performed that is of lower dimensionality than the desired information, or is sparse in at least one of the dimensions. This type of sparse imaging is investigated in this thesis.

Sparse imaging is typically employed in image guided interventions and surgeries, where high speed, high image resolution and an open acquisition setup are of major importance. The 3D position of the surgical instruments with respect to the 3D patient anatomy is then assessed through 2D imaging such as X-ray fluoroscopy [39, 101] or Ultra-sound [11, 127]. For similar reasons mono- and biplane X-ray fluoroscopy became the standard for kinematic analysis of joints, allowing the acquisition of a wide range of motions, such as running [4] and jumping [36, 122]. In other situations, radiation dose and cost reduction play a mayor role in employing sparse imaging. Bone surface reconstruction from points pin-pointed during knee surgery e.g. has been investigated for replacing prior CT acquisition [43, 104]. Also, assessment of organ motion, such as cardiac or respiratory motion, can occur from temporally sparse CT images [58, 92].

In all these cases reconstruction of the structure of interest in its native dimensionality is desirable. The next section describes the four medical applications of sparse imaging, which we investigate in this thesis, in more detail.
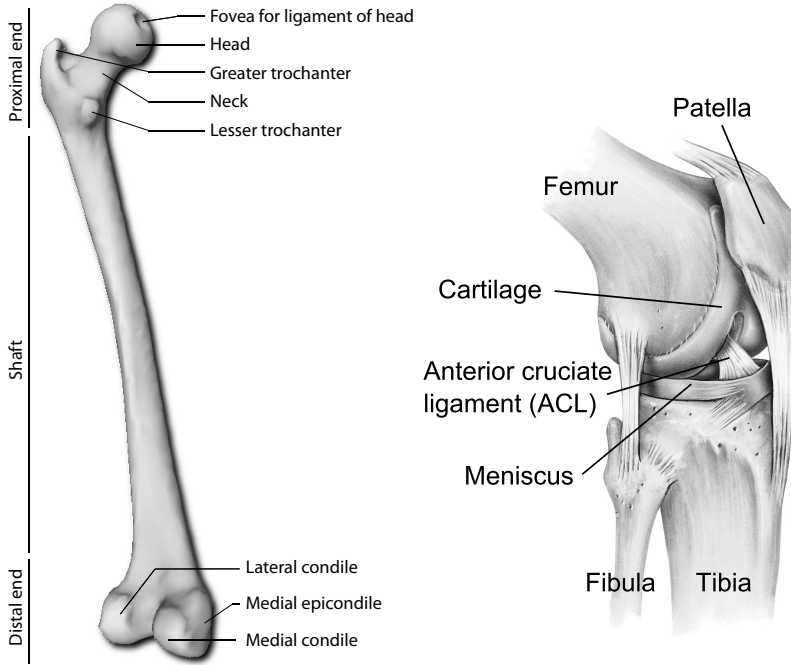
Figure 1.1: Left: Anatomy of the human femur from posterior (behind) view. Right: Anatomy of the human knee from anterolateral (frontal side) view.

1.2   Sparse imaging applications

We investigate sparse imaging in two medical fields. In the first part of the thesis the main focus lies on knee imaging for pre-operative planning and motion assessment. In the second part of the thesis we investigate the heart, and the coronary arteries. In this section we describe the medical applications, and the added value of shape or motion reconstruction in these cases.

1.2.1   Pre-operative planning for knee replacement surgery

Knee replacement surgery is the final solution to relieve pain in the knee due to carti-lage damage. A conventional surgery is planned using X-ray images of the knee or the entire leg. The components are placed into the knee in a standardized position, and implant size is determined during surgery using different sized trial components. Cor-rect positioning of the implants is important for proper functioning of the implanted knee [82]. Recently, optical guidance systems have been introduced to facilitate ac-curate component placement [53, 114]. Most recent guidance systems also enable patient specific surgery planning besides intra-operative navigation. These systems require a patient specific 3D model of the bones. In current practice these models

are obtained from a pre-operative CT scan, causing higher costs, and exposing the patient to a larger amount of radiation than the conventional method.

If the 3D bone surface could be reconstructed from a small number of 2D X-ray images, both radiation dose and costs would decrease.

We primarily focused on the femur (thigh bone) geometry, and its reconstruction from 2D projections. An illustration of the femur is given in Figure 1.1. The femur bears the weight of the upper body when standing, and accounts for approximately one quarter of the total body length [96]. Its distal (lower) side articulates with the tibia in the knee, and the proximal (upper) side articulates with the pelvic bone in the hip joint. The proximal femur is composed of a femoral head, which is connected through the femoral neck to the shaft, and two elevations, the greater and lesser trochanters serving as muscle attachment sites. The distal femur is composed of two condyles, medial (inner) and lateral (outer), which are horizontal in standing position.

1.2.2 Knee kinematics for ACL assessment

The knee (Figure 1.1) is the largest human joint, allowing flexion and extension of the leg. The knee joint is mechanically relatively weak. Its stability depends on the surrounding muscles and ligaments [96]. The anterior cruciate ligament (ACL) is one of the two crossing ligaments connecting the femur and the tibia, as shown in Figure 1.1. It restricts the motion of the tibia with respect to the femur, in particular the twisting motion and the front to back sliding motion. Non-contact injuries to this ligament occur frequently due to a sudden stopping or turning motion, and is most relevant in female athletes [18]. The ACL has to be surgically replaced or fixed after a tear.

If one could measure the chance of an ACL injury for an individual, a special training program could be constructed to strengthen surrounding muscles stabilizing the knee. Research is therefore performed to investigate the femoral and tibial kinematics during different motions. Joint kinematics is commonly measured using monoplane [36, 47] or biplane X-ray fluoroscopic sequences [4, 122, 135]. In this thesis calibrated biplane fluoroscopic sequences were available. An illustration of the acquisition setup is shown in Figure 1.2. Subjects were asked to jump from a 40 cm high box, and perform a stiff landing on one leg in the field of view of the biplane camera system [122]. The acquired fluoroscopic sequences were then analyzed for femoral and tibial motion. The current standard for obtaining accurate bone kinematics from fluoroscopic sequences is to register a patient specific 3D bone surface frame by frame. Patient specific bone surfaces are obtained by segmenting CT [4, 36, 122] or MRI [1] images.

If the patient specific 3D bone surface could be estimated from the 2D fluoroscopic images rather than from a CT or MR acquisition, that would lower the study costs, and in the CT case also the radiation dose. This is advantageous, as ionizing radiation is known to increase the chance for cancer. A further drawback of the current approach is the need for manual interaction in every fluoroscopic frame. To minimize this labor intensive process often only a subset of the frames (e.g. every 4th frame) is used to study the motion [122]. An automated tracking method could therefore
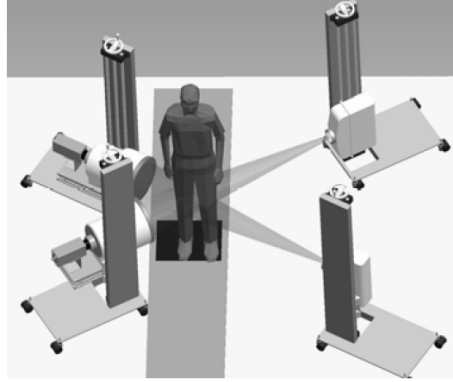
Figure 1.2: Biplane fluoroscopy setup used for acquiring the still images and sequences used in Chapter 2 and 3. The Figure is reproduced with permission from [122].

decrease the costs and the time requirements, while potentially providing kinematics of higher temporal resolution.

1.2.3   Personalized cardiac motion prediction from end diastolic CT

The heart periodically contracts to pump de-oxygenated blood into the lungs, and oxygen rich blood into the body. It consists of four chambers: right and left atria, and right and left ventricles, as shown in Figure 1.3. The ventricular contractions (systole) generate the main blood flow. The atria are the receiving chambers collecting the blood that will be sucked into the ventricles during their relaxation (diastole). The right side of the heart pumps blood into the lungs, and the left side into the body. Ventricular systole duration is about one third of the cardiac cycle at rest state. With increased heart rate, systole duration can occupy as much as half of the cardiac cycle [49].

Knowledge of the patient specific cardiac (heart) motion was shown to increase the robustness of 2D+t/3D+t registration of X-ray angiographic (XA) sequences and coronary arteries extracted from CT angiography (CTA) [90]. Such registration is performed to improve guidance during coronary interventions (see next section). However, current prospective gating technologies allow CT acquisition with zero dose for time intervals outside the window of interest (usually end-diastole) to decrease radiation dose. Cardiac motion information in these cases is not available. 4D scanning of selected patients who undergo a percutaneous coronary intervention is also not an option, as the decision on the treatment is made after the CT acquisition. As a consequence, patient specific cardiac motion is expected to be less frequently available for registration in the future.
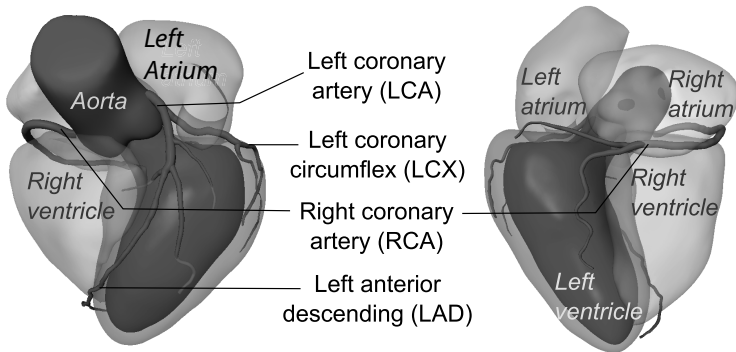
Figure 1.3 (color version on page 147): Illustration of the four heart chambers, the aorta, and the main coronary arteries. The coronary arteries directly branch off from the aorta, and provide blood to the cardiac muscles. Due to the transparency of the chambers both endo- and epi-cardium are visible for the left ventricle. The heart images were kindly provided by Hortense Kirisli.

Prediction of the patient specific cardiac motion from end-diastolic CTA scans could therefore provide the necessary temporal information for 2D+t/3D+t registration of XA sequences and 3D coronary arteries.

1.2.4   Image guidance during Percutaneous coronary intervention

The coronary arteries supply blood to the cardiac muscles, branching off directly from the aorta, as shown in Figure 1.3. During adult life, plaque is building up in the vessel wall of the coronary arteries causing narrowing of their diameter. If the plaque completely fills up the vessel (total occlusion), or if the plaque raptures and the floating part blocks the vessel, the heart muscles do not get enough blood, and myocardial infarction occurs.

To prevent the plaque from rupturing, and to open up the lumen, percutaneous coronary intervention (PCI) may be performed. During PCI a catheter is inserted through the femoral artery into the coronary artery. When the inserted guidewire passes the occlusion, a balloon is inflated to widen the lumen, and a stent is placed. Navigation through the arteries is facilitated by contrast enhanced X-ray angiography (XA) sequences. At planning time both XA sequences and CT angiography (CTA) scans are available to the interventional cardiologist. Intra-operative XA sequences are taken using a mobile X-ray system mounted on a C-shaped robotic arm, allowing change in view angle on demand. A typical system is depicted in Figure 1.4. For more challenging interventions biplane systems have been developed. These systems consist of two calibrated C-arms, making the 3D mental reconstruction of the arteries easier for the cardiologist. The XA sequences show the coronary arteries due to the injected contrast agent, the catheter and the guide wire. Lesions can be seen as narrowed vessel parts. In the case of a chronic total occlusion (CTO), the vessel is

Figure 1.4: A typical monoplane C-arm system with operating table, screens and control unit. The image shows the Philips Fixed II Screening Suite.

only partially visible, because the contrast agent is not able to pass the occlusion. This poses a difficulty in treatment, lowering success rates of PCI procedures from 99% for non-occluding lesions, to 60-80% for CTO procedures [37, 115]. On the other hand, the pre-operative CTA image contains both occluded and non-occluded arteries, with a possibility to distinguish calcified plaque [83].

Registration between XA sequences and end-diastolic CTA images would enable to overlay the CTA derived coronary artery and plaque information on the intra-operative XA data. Such improved guidance may therefore increase success rate for CTO procedures.

1.3   Statistical models complementing sparse information

In many situations when sparse imaging is performed, the dense reconstruction would enhance the value of the image data. A few examples have been listed in the previous section. The desired dense structure can be inferred by approximating the missing information. This is generally an ill-posed problem, but it can be restricted to plausible solutions using prior knowledge, e.g. in the form of smoothness restrictions [32, 48], or by employing statistical models representing plausible solutions learned from a training set. Depending on the nature of the desired information, statistical shape, motion, or deformation models have been proposed. The principles of these statistical models are very similar. We outline the shape model case below.

1.3.1   Statistical shape models

Statistical shape models are generative models able to generate unseen shapes based on a training population. If the training shapes are representative for the population,

the model spans the entire shape space, can thus generate all possible shape instances. Such shape models were first proposed by Cootes and Taylor [27].

Statistical models can be computed based on a set of shapes represented by corresponding features. Several representations have been proposed in literature, such as spherical harmonics [118], distances from the medial axis [103, 116], or positions of anatomically corresponding points, called landmarks, on the surface of the shapes. Point based models are most common due to their simplicity and generality, posing no restrictions on surface topology. Modeling of shapes in a point representation is termed point distribution models (PDM). In this thesis we employ point distribution models, and use the term statistical shape model in further chapters as a synonym for PDM.

The first step for creating a PDM is establishment of point correspondences between different shape instances. While manual correspondence assignment on 2D shapes may be feasible, automated methods are necessary for 3D or higher dimensional shapes. Correspondence methods in the literature may be categorized as pair-wise, propagating landmark points from a reference shape to all other shapes [42, 43, 56, 92], or group-wise, optimizing all shapes at once e.g. using minimum description length [33], the determinant of the covariance [70], maximizing both the geometric accuracy and the statistical simplicity [22], using a markov random field [100], or by a maximum a-posteriori framework [59]. Group-wise methods tend to produce more compact models [55], though pair-wise methods have also shown good performance.

The final step for creating a PDM is modeling of the shape distribution, as shown in Figure 1.5. To model shape variability rather than pose differences, usually every shape is aligned with Procrustes analysis including translation and rotation (optionally also isotropic scaling) to a reference space. Shapes are then represented by the vector of their concatenated landmark coordinates. These vectors can be regarded as a point in a high dimensional space, and accordingly, the set of shapes can be regarded as a point cloud. This point cloud is generally modeled as a multivariate Gaussian distribution, though non-linear models have also been proposed [31, 125]. We employ in this thesis a Gaussian model with a mean shape $\bar{s}$ and main modes of variation $\mathbf{\Phi} = [\phi_1, \phi_2, ...]$. These modes of variation are obtained using principal component analysis (PCA) on the aligned shapes, and represent the directions of highest variance. As a rule of thumb, 95 - 98% of the variance is kept during model construction, eliminating the smallest modes of variation likely to be noise. New instances can be generated with the model as

$$s = \bar{s} + \mathbf{\Phi}b \quad , \tag{1.1}$$

where b is the parameter vector defining the final shape s. All generated shape instances in such a model are a linear combination of the training shapes. As shapes far from the mean shape are less likely, models are generally either restricted with a hyper-ellipse to generate shapes within the 99% confidence interval, or are restricted with a hyper-cube not to exceed the ± 3 standard deviation bound (99.7% confidence limit) at any of the modes [31].

Aligned training shapes with point correspondence. Marks show corresponding points

Vector representation of the aligned shapes

Training set of shapes represented as a point cloud

Gray arrows show mean and two modes of variation. Ellipses show equally likely shapes in the shape space.
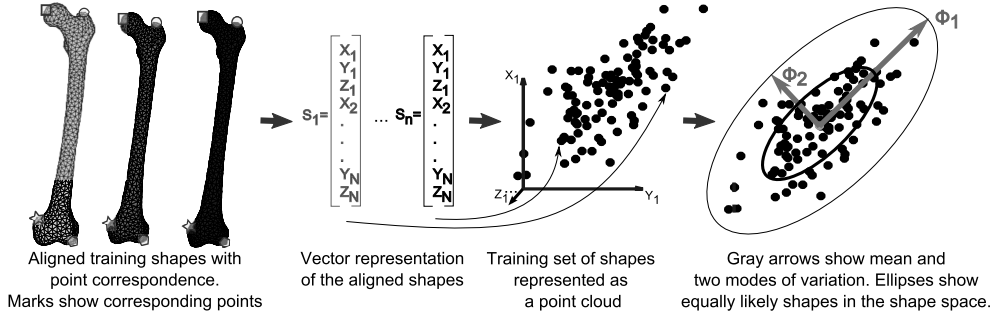
Figure 1.5: Illustration of the statistical shape model creation with a point distribution model.

Note, that the above methodology does not restrict the dimensionality of the shape. Shapes in 2D, 3D, and 4D (3D + time) can be modeled in the same manner. When modeling 4D shapes or motion the corresponding landmark points (or motions) are aligned to a reference space both spatially and temporally. Due to these similarities we use the term statistical shape model in its general meaning without restrictions on dimensionality.

We refer to [55] for a thorough overview of statistical shape models, and their various application areas.

1.3.2   SSM Fitting to sparse information

Statistical shape models are well suited for dense shape reconstruction from sparse information, as they restrict the final solution to lie within the space of plausible shapes. The common approach for shape reconstruction involves the estimation of the pose and the shape parameters which minimize a distance function quantifying the model's distance from the known information [11, 38, 44, 73, 109, 134, 142], often while keeping the shape estimate within plausible limits [7, 12, 45, 60, 104, 137, 141]. The actual choice of distance metric depends on the task, and on the sparse information.

The sparse information may consist of points with known corresponding counterparts in the statistical shape model. This scenario occurs when predicting a shape based on neighboring shapes [34, 62, 77, 106, 133], when the sparse information is represented by the same landmarks as the training set (Chapter 4,[88]), or by first estimating a fixed point correspondence between data and SSM [104, 139]. In this scenario a straight forward choice for the distance measure is the mean squared distance of corresponding points. Due to the knowledge of a well defined subspace of the training shape space, in this special case analytical solutions to the reconstruction problem are possible. These analytical techniques are regressions. To minimize the chance of over-fitting to the high dimensional small sample size training set, mostly linear regression methods have been proposed, such as principal component regression

[14, 15, 62, 34], ridge regression [16, 104, 139], and partial least squares [15, 106, 133].

Often, correspondences between the sparse information and the SSM are not known. This is e.g. the case in 2D-3D reconstructions from X-ray or ultrasound images. For these situations two main types of distance metrics have been proposed in the literature. One type samples a dense correspondence between image and model, and the other a sparse correspondence. We discuss the difference between these two types for the projection case.

Dense correspondence based reconstructions are most often based on the image intensity (though gradient based 2D-3D distance metrics have also been proposed [79, 121]). Model matching is performed by minimizing the difference between the image intensities and the intensity values produced with the model by projecting it to the X-ray plane. These intensity values are calculated by digitally reconstructed radiograph (DRR) generation, which simulates X-ray acquisition with attenuation calculated from the intensities inside the model. In this case thus beside the shape itself, knowledge about the intensity inside the shape is also required. The inside intensities can be derived from a reference image [60], can be synthetically created [119], or can be derived from the statistics of a training set [132, 134, 137]. The actual 2D/3D distance metric optimized varies in literature, examples include mean square intensity difference [132], mutual information [134], normalized mutual information [109], gradient difference [60] or gradient correlation [119].

Sparse correspondence based reconstructions minimize a distance metric quantifying the difference between model surface or projected model silhouette and their corresponding points in the image. Corresponding points are selected from salient image features. These are typically thresholded filter results, such as object edges [7, 6, 12, 38, 44, 72, 73, 141] or centerlines [40, 67]. The employed distance metric is based on the Euclidean distance between corresponding point pairs either in 2D [6, 12, 73, 38] or in 3D [7, 44, 139], but can also include other features e.g. gradient orientation [7, 6, 38, 40]. These methods are computationally less expensive than dense methods, as only the geometry of the shape is considered, which also simplifies model creation, as inside intensities and tissues are not included in the model. However, finding the feature points in the image corresponding to the model may be challenging due to numerous false findings, or missing correspondences due to low contrast. Consequently, these methods generally require the model to be placed close to the optimum [12, 73], or use a manual or semi-automatic edge/point selection [38, 44, 72, 140, 141]. For a complete review of 3D/2D registration methods we refer to [84].

1.4 Thesis contributions and overview

The goal of this thesis was to investigate the applicability of statistical shape models to complement sparse imaging data, and to develop new generic techniques for automatically fitting them to sparse data. We dealt with two major sparse reconstruction scenarios, namely 2D/3D reconstruction, which we successfully extended to incorporate a temporal dimension, and regression based shape reconstruction, where

Anatomy	Reconstruction	
	2D/3D	*Regression*
Femur	2, 3	6
Heart		4, 5, 6, 7
Coronary	7	7

Table 1.1: Chapters organized based on treated anatomy and reconstruction type.

Statistical analysis on	Reconstruction	
	2D/3D	*Regression*
Shape	2, 3	4, 5, 6, 7
Motion	7	4, 7

Table 1.2: Chapters organized based on target of statistics and reconstruction type.

correspondences between sparse data and model are known. We applied and validated the proposed methods for several medical applications on three main anatomies: the femur, the heart, and the coronary arteries. An overview of the chapters regarding the sparse reconstruction type and treated anatomy can be seen in Table 1.1. Another possibility to categorize our work is based on the structure statistics was performed on. In this sense we have modeled shapes and their motion. Table 1.2 shows the Chapter overview in this categorization.

The main contributions of the chapters are summarized below:

- In **Chapter 2** we propose a feature based 2D-3D reconstruction method for the femur from biplane X-ray radiographs. The potential application for this work could be a pre-operative planning tool for knee surgery. A novel edge based distance metric is developed for the reconstruction, which takes image gradient directions into account, and which is robust to non-orthogonal acquisition setups. Image gradients help to distinguish between parallel edges with opposite directions, as often occurs at joint surfaces, and non-orthogonal acquisitions are typical for e.g. Roentgen stereophotogrammetric analysis (RSA) setups. To overcome the small convergence radius problem of edge based reconstructions, an annealing technique is used in the optimization.

- In **Chapter 3** we extend the work presented in Chapter 2 to enable kinematic analysis of joints in 2D+t fluoroscopic sequences. By reconstructing the bone shape from the fluoroscopic sequences, we eliminate the need for a prior CT acquisition. We propose to reconstruct the bone shape based on all frames of the sequence, and show that this approach outperforms single time-frame reconstructions in both accuracy and convergence. Furthermore, we extend the feature based distance measure with appearance information to enlarge the convergence interval of the matching. This results in a method that converges in more than 90% of the frames using solely one manual pose per sequence, and without the need of annealing techniques in the optimization.

- In **Chapter 4** we evaluate the extent to which cardiac motion can be estimated from a 3D end-diastolic cardiac shape. This is an example of known point correspondences between the sparse data and the statistical model, thus allows an analytic solution to the reconstruction problem. We propose to use principal component regression, and show that the average prediction accuracy is about 1 mm point-to-surface distance, and is dependent on the cardiac phase.

- In **Chapter 5** we focus on the role of uncertainty of the sparse data in the case of known point correspondences between sparse data and the model. We propose a maximum a-posteriori ridge regression scheme, which takes the data uncertainty into account. Besides the final prediction accuracy, we also investigate the confidence of the resulting shape estimate. Our results showed an improved shape estimation with the proposed method compared to regular ridge regression.

- In **Chapter 6** we further investigate the role of uncertainty on shape estimation with known correspondence. We specifically investigate under what conditions it is advisable to incorporate uncertainties in the estimation. For this purpose we compare several methods for regression based shape estimation, including standard methods, methods less widely applied, and the method of Chapter 5. We found that incorporating uncertainty mainly improved performance in case training set and test data have considerably different uncertainties.

- In **Chapter 7** we combine the 2D/3D non-rigid registration and the 3D/4D reconstruction work presented in the previous chapters to perform 2D+t/3D registration of coronary arteries. We register XA sequences taken during intervention with pre-operative 3D end-diastolic CTA images. We propose to guide and constrain the temporal deformations with the use of motion models. Accordingly, we propose a methodology for creating population based coronary artery motion models and estimates such as the average coronary motion of the population. We evaluate the 2D+t/3D registration accuracy using several motion models and estimates. We found that the registration accuracy with the motion models was comparable, if not better than the accuracy with the 4D CTA based motion estimate.

This thesis condenses the work performed by the author during her PhD research. Every Chapter is based on a peer-reviewed journal or conference paper, therefore, some overlap in the chapters is inevitable. The further chapters describe the work in more detail.

Chapter 2

2D-3D reconstruction of the distal femur from biplane X-ray imaging using statistical shape models

Abstract

Three dimensional patient specific bone models are required in a range of medical applications, such as pre-operative surgery planning, improved guidance during surgery, modeling, simulation, and in-vivo bone motion tracking. Shape reconstruction from a small number of X-ray images is desired as it lowers both the acquisition costs and the radiation dose compared to CT. We propose a method for pose estimation and shape reconstruction of 3D bone surfaces from two (or more) calibrated X-ray images using a statistical shape model (SSM). User interaction is limited to manual initialization of the mean shape. The proposed method combines a 3D distance based objective function with automatic edge selection on a Canny edge map. Landmark-edge correspondences are weighted based on the orientation difference of the projected silhouette and the corresponding image edge.

The method is evaluated on the reconstruction of the femur bone from biplane cadaver X-rays, in-vivo fluoroscopic images, and an in-vivo fluoroscopic sequence of the knee. For comparison we also registered the ground truth shapes, segmented from CT images, with the same data. We report the point-to-surface (P2S) distances to the ground truth shape in a reference pose, and the remaining distance from manually selected femoral edges in the image.

Rigid registration of the ground truth shape to the biplane fluoroscopy achieved sub-millimeter accuracy (0.68 mm) measured as the root mean squared (RMS) P2S distance. The non-rigid reconstruction from the biplane fluoroscopy using the SSM also showed promising results (1.68 mm RMS P2S). A feasibility study on one fluoroscopic time series illustrates the potential of the method for motion and shape estimation from fluoroscopic sequences with minimal user interaction.

2.1 Introduction

Three dimensional pose estimation and shape reconstruction from a sparse set of 2D projection images is important for a range of orthopedic applications. Computer navigation systems based on patient specific bone models may improve pre-operative planning [71], and yield higher surgical accuracy [53, 114]. Moreover, such models are a prerequisite for personalized finite element simulations, which may improve implant designs, and may provide real-time feedback during implant placement [52, 71, 85]. Also, personalized 3D bone models enhance the accuracy of in vivo motion tracking, that is used for studying the relation between anatomy, motion and disease [36, 122].

Pose estimation from biplane imaging in case of a priori known shape has been widely studied in the literature. We will not discuss this case here, and refer to a recent review on existing techniques [84]. Such prior 3D shapes are often reconstructed from CT scans. As these scans are associated with relatively high costs and radiation doses, investigating methods for bone reconstruction from a small number of low-dose, low-cost X-ray images is relevant.

Shape reconstruction from sparse projection data requires prior information to guide the reconstruction process, which can be provided by integration of expert knowledge [74], or by Statistical Shape Models (SSM) [12, 60, 72, 73, 75, 119, 134, 139]. The latter methods are either based on intensity or on edge features in the images.

Intensity based methods optimize intensity similarity between the X-ray images and digitally reconstructed radiographs (DRRs). Yao et al. built a statistical model from tetrahedral meshes, and assigned an analytic density function to every tetrahedron used for DRR generation [134]. Tang et al. proposed a hybrid global-local shape model, and "grew" the surface inward to simulate thickness of the cortical bone of femora [119]. Hurvitz et al. proposed 3D fitting by non-rigid registration of X-ray and DRR, using a statistical deformation model as prior, a template CT-like intensity image for DRR generation, and a template bone surface as shape representation [60].

Edge based methods minimize the distance between object edges in the image and model silhouette. Fleute et al. [44] presented an extension of the rigid registration method of Lavallee et al. [75] to deformable models minimizing the 3D back-projected edge to surface distances for both shape and pose-parameters. Zheng et al. suggested to determine point correspondences for the fitting after rigid alignment, and do least squares shape estimation using the SSM [139]. Other approaches did not explicitly build point correspondences, thus prevent errors that might occur due to inaccurate or wrong point correspondences. Kurazume et al. [72] proposed a method that minimizes the sum of 2D distances of the image contour and the projected outer silhouette after rigid alignment of the mean shape by the method of Iwashita et al. [61]. Another correspondence free approach, by Benameur et al., simultaneously optimizes a prior term and a 2D error metric after an initial rigid registration based on a number of anatomical landmarks [12]. The data fit term in their work was based on an edge potential field, computed as the Gaussian blurred edge map of the X-ray images. The average of the edge potential field sampled at the projected contour pixels was subsequently maximized. The approach was validated on 3D vertebrae reconstruction. Lamecker et al. proposed a similar energy computation with the projected

outer silhouette only, solving in this way the inner contour correspondence problems hampering reconstruction of the pelvic bone [73]. To achieve a better correspondence [40] and [38] include the normal orientation in the error metric.

The algorithms discussed above can be categorized as either 2D similarity based [12, 38, 72, 73, 119, 134] or 3D similarity based [44, 60, 139]. A 2D similarity metric enables automatic edge selection [12, 73] and correspondence free matching [12, 38, 72, 73, 134, 119], but is prone to perspective projection distortion amplifying the effect of object surfaces closer to the X-ray source and further away from the projection plane. Three dimensional similarity based techniques are therefore especially favored in a non-orthogonal imaging setup.

Another categorization can be made on whether a method requires the delineation of the complete (or partial) silhouette of the object on the X-ray images prior to matching [38, 44, 61, 65, 72, 75, 139, 141], or not [12, 60, 73, 119, 134]. Manual contour drawing on a clinical X-ray may be inaccurate and cumbersome, and automation is not trivial due to the overlapping structures. Most approaches either do not discuss the edge delineation problem from clinical X-rays [38, 44, 72, 75], or use manual or semi-automatic contours [65, 139, 141]. Methods that do not need delineation are either intensity based [60, 119, 134], or implement automatic edge selection based on a 2D similarity metric [12, 73]. We propose a new edge selection based method because of the relatively well visible edges of the femoral bone in the X-ray images, and the lower computational complexity of the edge-based methods compared to DRR based methods.

The goal of this work is reconstruction of the 3D patient specific femur from two (or more) calibrated X-ray images taken from different orientations. The contributions of this chapter are three-fold:

- The proposed method combines the benefits of a 3D similarity metric [44, 139] with an automatic edge selection scheme [12, 73].

- An orientation based correspondence weighting (an extension of [63]) provides robustness w.r.t. missing, incorrect and noise edges in fluoroscopic images.

- Additionally, in contrast to current methods that only model the part of the bone in the Field of View (FOV) e.g. proximal or distal part of the femur, we use a statistical model of the complete femur for fitting to knee X-ray images, exploiting all information present in the FOV.

This results in a novel automated reconstruction method requiring a single manual initialization. In case of reconstruction of a temporal sequence, initialization may come from the previous frame, such that user interaction is further minimized. Our method builds on the work of Benameur et al. [12] in that we optimize shape and pose of an SSM based on the Canny edge map of the X-ray images. Our approach differs, however, in that we optimize the 3D error of the SSM (as opposed to a 2D error of the projected contour pixels) and include orientation weighting in our framework. Though normal orientation has been introduced before in 3D-2D non-rigid registration, the way it is incorporated in our work differs from previous methods.

Feldmar et al. extended the spatial dimensions of edge points with the dimensions of the normal vector, such that distance and orientation were handled equally [40], whereas in our approach normal orientation is a weighing factor suppressing unlikely correspondences. The work of Dworzak et al. [38] is more similar to ours, however, the proposed error metric is different in that our objective function has an intrinsic robustness to outliers. We thoroughly validate the method on the distal femur using cadavers, in-vivo biplane fluoroscopic images, and an in-vivo biplane fluoroscopic sequence with known ground truth. We also investigate the effect of different algorithm parameters on the reconstruction accuracy.

2.2 Method

2.2.1 Method Overview

Our method is based on the optimization of an objective function E that contains both a data fitting term E_{fit} and a shape prior term E_{prior}:

$$E = E_{fit} + \lambda E_{prior} \quad , \tag{2.1}$$

where λ is a weighting factor. The prior knowledge of plausible shapes is described by a point distribution model [27] as explained in Section 2.2.2, and the fitting term, explained in Section 2.2.3, describes how well the shape fits the edges in the X-ray images. Figure 2.2 shows a graphical overview of the method.

2.2.2 Statistical Shape Model and Shape Prior

The point distribution model is built from N training shapes segmented from CT data. A correspondence algorithm is applied to represent each shape with the same number n of corresponding landmark points p_j, $j = 1..n$ [42], which are interconnected to form triangulated surfaces. Principal component analysis (PCA) of the aligned shapes results in a model of the form

$$x \approx \bar{x} + \Phi b_x \ , \tag{2.2}$$

where $\bar{x}$ is the mean shape, Φ is the matrix of the main modes of variation, and b_x is the representation of the shape x in parameter space.

To penalize improbable shapes during optimization, the Mahalanobis distance of the shape to the model mean is utilized as the shape prior in the objective function:

$$E_{prior} = b_x^T \Sigma^{-1} b_x \quad , \tag{2.3}$$

where Σ represents the covariance matrix of the shapes in the parameter space.

2.2.3 Data Fit Term

The data fit term E_{fit} is a function that has low value when the current 3D shape accurately matches the edges in the X-ray images. Edges are computed prior to

matching using a Canny edge detector. We propose to assign a matching score to every landmark point, and average over all landmark points and over all projection directions to obtain the data fit function:

$$E_{fit} = \frac{1}{n * n_{proj}} \sum_{i=1}^{n_{proj}} \sum_{j=1}^{n} \Theta_i(p_j) \quad , \tag{2.4}$$

where n_{proj} is the number of projections, and n is the number of landmark points of the model. The matching score $\Theta_i(p_j)$ of landmark p_j in projection i is given as

$$\Theta_i(p_j) = 1 - u_i(p_j) * w_i(p_j) * D_i(p_j) \quad , \tag{2.5}$$

where the term $u_i(p_j)$ masks the silhouette landmark points such that it results in 1 for silhouette landmarks who's projection falls in the FOV of projection i, and in 0 otherwise. Silhouette landmarks were defined as the landmark points shared by two triangles with one normal pointing toward and the other pointing away from the source. The distance term $D_i(p_j)$ is defined as

$$D_i(p_j) = \exp\left(-\frac{d_i(p_j)}{\nu}\right) \quad , \tag{2.6}$$

where $d_i(p_j)$ denotes the 3D distance calculated as the distance of the landmark point p_j from the ray connecting the nearest Canny edge point of the projected landmark with the X-ray source[1]. The distance weighting factor ν controls the smoothness of the exponential. The distance term was constructed such that the 3D distances are mapped on the interval [0,1]. The inherent non-linearity of the function provides outlier suppression by converging to zero for large distances.

The orientation weighting term $w_i(p_j) \in [0,1]$ is defined as

$$w = \begin{cases} 0 & , \text{if } \alpha > 90° \\ \cos\alpha & , \text{otherwise} \end{cases} \tag{2.7}$$

where α is the angle between the projected normal vector of the 3D surface at landmark point p_j and the gradient of the image at the corresponding edge location. The closest point correspondence on the Canny edge map doesn't ensure correspondence of a landmark point to a relevant edge (a true object edge in the Canny edge map, in our case femur edges). Therefore, the orientation weighing down-weighs correspondences with large angle difference, minimizing the effect of probably false correspondences. The orientation weighing furthermore provides additional information about the landmark neighborhood, aiding convergence. We prefer to use the image gradient of the X-ray for the orientation weighing rather than the edge tangent of the edge map as in [63], as it allows differentiation between edges of positive and negative orientation,

---

[1]In case multiple landmark points map to the same point of the Canny edge map, only the correspondence of the closest landmark is kept, the other landmarks are masked out with the function u_i.

and thus it allows distinguishing parallel neighboring structures with similar shapes (such as joint surfaces).

Due to the subtraction in Equation 2.5, the optimal matching score is zero. Figure 2.1 shows $\Theta_i(p_j)$ as a function of the angle α and the distance $d_i(p_j)$. A step-by-step illustration of the proposed method is given in Figure 2.2.
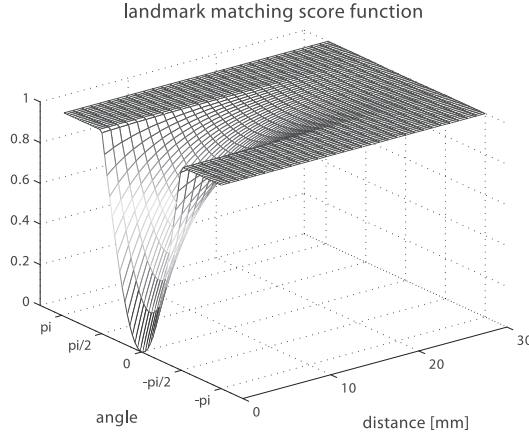
Figure 2.1: The silhouette landmark matching score $\Theta_i(p_j)$ with $\nu = 5$ as a function of the angle and the 3D distance. The non-linearity amplifies the effect of close points as well as small angles, and suppresses the effect of far points likely to be outliers, and large angles which probably constitute false correspondences.

2.2.4 Minimization algorithm

The objective function in Equation 2.1 has two sets of parameters: the pose parameters (translation, rotation, and isotropic scaling), and the shape parameters (the vector b from the shape model). The minimization strategy involves two stages: pose optimization of the mean shape, followed by simultaneous pose and shape estimation.

Pose optimization consists of translation followed by optimization of rigid pose and isotropic scaling. Due to the large amount of local minima, a number n_{opt} of affine optimizations are performed by adding small random displacements to the pose with lowest objective function value.

All optimizations were performed using preconditioned conjugate gradients in a trust region approach, as implemented in the Optimization Toolbox of Matlab. The algorithm is guaranteed to terminate by applying thresholds on step size, gradient magnitude, and maximum number of iterations.
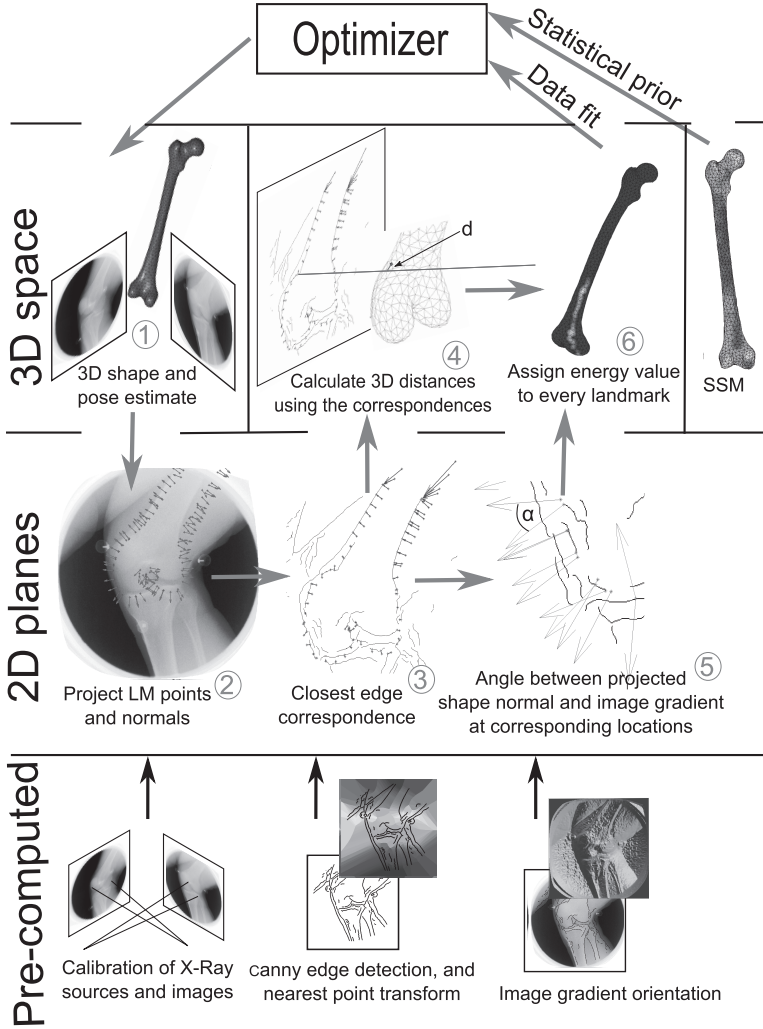
Figure 2.2 (color version on page 149): Method overview: The objective function is a combination of a shape prior term and a data fit term. The data fit term is calculated in six steps: 1) The silhouette landmark points of the current model are determined, and 2) projected to the plane; 3) The projected landmarks are matched with the closest edge point in the Canny edge map using a pre-computed nearest neighbor transform; 4) The 3D distance d between the silhouette landmarks and the ray connecting their corresponding 2D edge point with the X-ray source is defined; 5) The angle between the projected 3D surface normals at the silhouette landmarks and the gradient of the image at the corresponding edge positions are calculated; and 6) The matching score $\Theta_i(p_j)$ in Equation 2.5 is calculated for every landmark.

2.3 Data

Femur shape model   The femur shape model was constructed using a semi-automatic snake-based segmentation from CT Angiography scans of the lower extremities of 30 patients and CT scans of 13 cadavers.[2] The 43 segmentations consisted of 25 mirrored right, and 18 left femora. The population contained 30 male and 13 female subjects aged between 23 and 92 years. The bones were scanned on different scanners with voxel sizes ranging from 0.625 x 0.625 x 0.8 mm^3 to 0.782 x 0.782 x 0.8 mm^3. Landmark points on the training shapes were calculated using the GAMES algorithm [42] with an average landmark distance of 9 mm, yielding 931 landmarks for the entire femur. Retaining 95% variance at model construction resulted in 30 out of 42 modes of variation.

Cadaver data   Uni-planar biplane X-rays of two naked cadaver femora of male subjects (74 and 86 years old) with 1 mm tantalum markers were acquired using a standard Roentgen Stereophotogrammetric Analysis (RSA) setup. The cadaver femora were placed above a calibration box (Leiden Carbonbox 8), which enabled the computation of the exact acquisition geometry [126]. Two X-ray views of the femora were acquired simultaneously, with an angle of 35° , containing approximately 50-60% of an average length femur. For computational reasons, each X-ray image was down-sampled to a pixel size of 0.47 x 0.47 mm. Each femur was imaged in four orientations: anterior-posterior (AP), medial-lateral (ML), lateral-medial(LM), and posterior-anterior (PA). The AP orientation was scanned twice with a small displacement. Ground truth pose and shape were obtained from CT scans of the bones including the markers.

Fluoroscopic data   We used subset of the data of an ongoing study designed to give insight in 3D knee kinematics during jump-landing (e.g., [122]). Subjects were asked to jump from a 40 cm high box to the ground in the field of view of a custom built biplane fluoroscopy system. The system was focused on the dominant landing support leg, and consisted of two modified C-arms (Philips Medical systems, Best, NL) placed at approx. 80 degrees and equipped with synchronized high-speed cameras recording at 500 frames/second with a resolution of 1024x1024 pixels. For further information on the calibration we refer to [122]. Biplane fluoroscopic jump sequences and CT scans were acquired for 10 healthy subjects (9 male and 1 female, aged 23-41 years). The FOV of the CT ranged from approximately 12 cm above to 12cm below the joint line of the knee.

   We used single time-point biplane knee images of the 10 subjects, as well as the complete fluoroscopic jump sequence of one subject. Triangulated CT segmentations (extracted with Mimics, Materialize Inc., Plymouth, MI) were used as ground truth shapes, and ground truth pose (translation, rotation and isotropic scaling) was derived by 3D-2D registration of the ground truth shape to manually selected canny edges of the fluoroscopic images, using commercially available software for Model-Based

---

[2]All scans were performed for other reasons than osteoarthritis.

Roentgen Stereophotogrammetric Analysis (MBRSA) (Medis-specials, Leiden, The Netherlands)(for details see [65]). For the fluoroscopic jump sequence, ground truth pose was calculated in every fourth time-frame. We note, that because of the high-speed cameras the fluoroscopic data used in this chapter is lower quality than standard fluoroscopy.

None of the subjects of the fluoroscopic data was included in the training data for the femur shape model.

2.4   Evaluation Scheme

2.4.1   Evaluation methodology

We adapted the evaluation methodology proposed by [129], i.e. reconstructions were performed with initializations of varying distance from the ground truth (measured in point-to-surface distance, see Section 2.4.2). Initializations were derived by random displacements (translation, rotation, scaling) of the mean shape in the ground truth pose, such that the initial accuracies were distributed within the distance range $[P2S_L,14]$ mm, where $P2S_L$ is the shape distance between the model mean and the target shape. To ensure an even distribution throughout this range, seven equally sized sub-intervals were used. Maximum displacements per parameter per interval were found via line search. Additionally, parameters were constrained to the range [-10,10] mm translation, [-10,10] degree rotation, and [90,110] % scaling. This was necessary due to the elongated, tubular nature of the femoral bone. Displacements were selected randomly from a uniform distribution within the calculated parameter maximum bounds, and used as initialization if the resulting shape initialization belonged to the desired interval.

2.4.2   Error metrics

Performance was evaluated with three types of distance metrics: point-to-surface distances, edge-surface distances and absolute pose differences.

The point-to-Surface (P2S) distance of the shape estimate to the ground truth shape was used to evaluate the 3D reconstruction accuracy. P2S errors were only evaluated at the landmark points of the knee.[3] The P2S distance of a shape is given as the root mean squared (RMS) distances of its landmark points. Matching convergence was defined as an RMS P2S error smaller than 3 mm.

The edge-Surface (ES) distance was employed to indicate how well the object edges[4] in the projection images were approximated by the 3D model. It was calculated as the 3D distance between the ray connecting the selected edge pixels with the X-ray source, and the fitted 3D surface. The edge-surface distance of an entire shape was given as the root mean squared (RMS) ES distance of all selected edge pixels.

---

[3]Up till approximately 12cm above joint line.

[4]In the fluoroscopic data object edges were defined as the manually selected Canny edges from the ground truth calculation (see Section 2.3). In the cadaver experiments object edges were defined as the Canny edges close to the projected ground truth shape.

Absolute pose differences were computed to determine the pose error between ground truth and reconstructed pose per parameter (translation, rotation, isotropic scaling). Translation parameters were based on world coordinates, while rotation and scaling was performed in the shape coordinate system with axes pointing in the Lateral-Medial, Anterior-Posterior, and Proximal-Distal directions. Despite the different parameters not being independent from each other, this metric does provide relevant information about rigid fitting accuracy.

2.4.3 Parameter settings

The proposed method incorporates four groups of parameters that have to be selected prior to application. All experiments in Section 2.5 were computed with the following values.

For the preprocessing of the X-ray images, we used $\sigma = 4$ pixels for the Gaussian blurring (matching the protocol of the ground truth pose generation of the fluoroscopic data), and an upper threshold of 0.04 for the Canny edge detection. For the lower threshold we chose a value of 20% of the upper threshold, which we found to be a sensible comprise between the amount of edges and computation time. The effect of varying these parameters on the performance of the algorithm is evaluated in Section 2.5.5.

The fitting term has a distance weighting factor $\nu = 5$, which is equivalent to an outlier distance threshold of approximately 20 mm. Note that different thresholds can be selected for a multi-resolution approach.

The optimization algorithm depends on the number and magnitude of random displacements during rigid optimization. We drew $n_{opt} = 5$ random displacements per rigid optimization from a normal distribution with a standard deviation of 3 (mm,° , and % respectively). These numbers were determined by pilot experiments without extensive optimization, and represent a trade-off between accuracy and computation time.

The objective function has a factor λ, which determines the relative weight of the prior term. A high value of λ results in a rigid model, which is able to fit the image edges less accurately, while low value of λ might result in implausible shapes. We optimized this parameter ($\lambda = 0.01$) on fluoroscopic subject nr. 1 only, to obtain unbiased results. Results showed that subject nr. 1 did not outperform other subjects, and is therefore not treated separately from the test set.

2.5 Experiments and Results

We performed five experiments to evaluate different aspects of the proposed method.

2.5.1 Rigid 3D-2D registration to fluoroscopy

Experiment The first experiment was designed to evaluate the data fit term of the objective function, eliminating all possible errors caused by the SSM. We performed

rigid 3D-2D registration of the ground truth shape to the biplane fluoroscopic image pairs of 10 volunteers with initial pose according to Section 2.4.1.

Results   A summary of the results is given in Table 2.2. Figure 2.3 gives a more detailed illustration of the achieved convergence radius and accuracy. The scatter plot (Figure 2.3 left) shows all fitting runs, while the box-plots (Figure 2.3 middle, right) include only the converged cases, where convergence was defined as a RMS P2S error smaller than 3 mm. We compared the quality of the pose estimation with the ground truth based on the edge-surface distances. Results in Figure 2.3 show a comparable rigid fitting accuracy of the proposed method and model based RSA used to extract the ground truth pose. The reconstructed pose varied with different starting positions and due to randomness involved in the optimization. The spread of fitting accuracies, however, remained small. Table 2.3 shows the absolute pose difference of the registration and the ground truth. The 95% convergence radius was 6.34 mm.

2.5.2   Non-rigid 3D-2D registration of Cadaver femora

Experiment   The second experiment was designed to test the non-rigid fitting on biplane X-rays of naked cadaver femora, i.e. in the absence of overlapping structures. This way, the effect of the Canny parameter selection was minimal, and the edge map contained nearly the entire silhouette of the object with a minimal amount of irrelevant edges. We used the cadaver data described in Section 2.3. The cadavers used for testing were excluded from the training set from which the SSM was built.

Results   The 3D-2D fitting of the SSM resulted in an average RMS P2S distance of 1.68 mm for all converged cases. Detailed results are shown in Table 2.1. In all test cases the accuracy of the rigid fit of the mean shape improved substantially with non-rigid fitting (on average by 0.47 mm P2S, and by 0.71 mm ES). An example biplane image pair in AP orientation is shown in Figure 2.4 with initial position, rigid and non-rigid fitting results. As a reference, a 3D-3D fitting of the SSM to the cadaver knee was also performed by minimizing the P2S distance between the model and the target shape. The resulting RMS P2S distance (1.1 mm for Cadaver 1, and 0.9 mm for Cadaver 2) represents the best possible reconstruction accuracy with our model.

2.5.3   Non-rigid 3D-2D registration to fluoroscopy

Experiment   In the third experiment we evaluated the performance of our complete non-rigid 3D-2D registration framework in a realistic setup. The statistical model of the femur was fitted to 10 fluoroscopic biplane knee X-rays using the proposed method. Separately, we assessed the effort of manual initialization starting from the mean shape in the middle of the FOV. The user could adjust the pose per parameter seeing the updated silhouette overlaid on the X-rays.
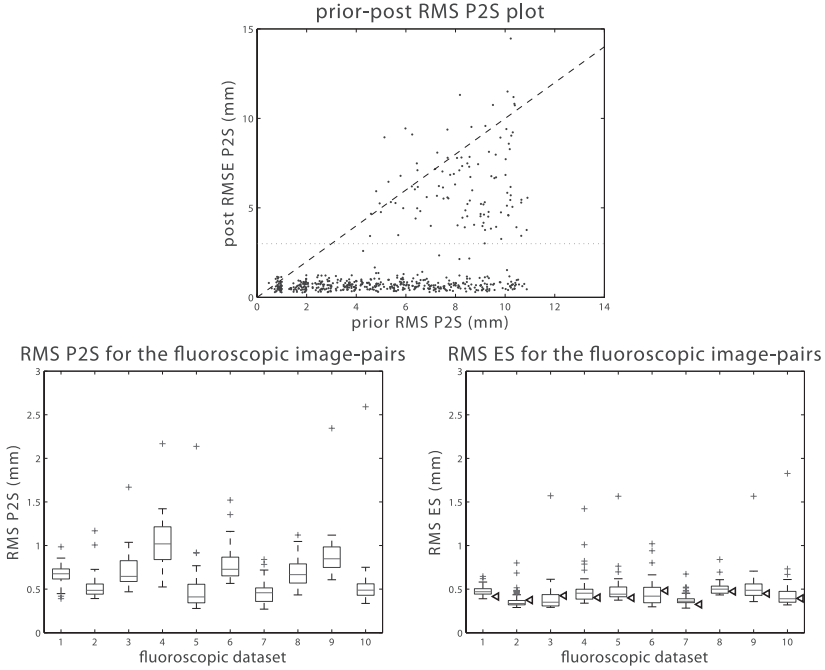
Figure 2.3: Rigid fitting of the ground truth shape on 10 fluoroscopic image pairs. Top: Convergence plot. Each dot represents a separate fitting with the initial P2S distance on the 'x' axis, and the post-registration distance on the 'y' axis. The horizontal dotted line shows the convergence threshold, the dashed line is the no improvement line. Bottom left: Box-plot of P2S distances per fluoroscopic dataset for all converged cases, Bottom right: Box-plot of ES distances per fluoroscopic dataset for all converged cases, the black triangles mark the ES distance of the ground truth pose.

Results A summary of the results is given in Table 2.2. The 3D-2D registration attained a mean P2S error of 1.33 mm and a mean RMS P2S error of 1.68 mm from the ground truth. To interpret these numbers, the optimal fit of the SSM to the 3D ground truth shape was computed by minimizing the P2S distance, and was found to be an RMS P2S distance of 0.78 mm on average. Convergence interval, and accuracy plots are shown in Figure 2.5. The 95% convergence radius was 6.76 mm. Pose errors of the reconstructions are listed in Table 2.3. An example fit is given in Figure 2.6. Supplemental movie 1 shows all stages of the fitting process for one subject[5]. Manual initialization with a few clicks per parameter resulted in RMS P2S distances smaller than 5 mm in all 10 cases.

---

[5]http://bigr.nl/movies/nbaka/movie1.avi

Table 2.1: Cadaver experiment results of non-rigid fitting of the SSM to the biplane X-rays of two cadavers in four scanning orientations. Mean RMS distances and standard deviations of the fitted SSM are given for all converged cases. Point-to-surface (P2S), and edge to surface (ES) distances are reported in mm.

#	Cadaver orientation	P2S	ES
1	AP a	1.78 ± 0.13	1.12 ± 0.26
1	AP b	1.81 ± 0.21	1.20 ± 0.33
1	PA	1.93 ± 0.19	1.50 ± 0.30
1	LM	1.89 ± 0.21	1.04 ± 0.08
1	ML	1.83 ± 0.27	0.96 ± 0.26
2	AP a	1.22 ± 0.10	1.20 ± 0.05
2	AP b	1.31 ± 0.15	1.29 ± 0.21
2	PA	1.77 ± 0.21	1.24 ± 0.09
2	LM	1.53 ± 0.09	1.38 ± 0.06
2	ML	1.73 ± 0.18	1.17 ± 0.09
	all	1.68 ± 0.29	1.21 ± 0.25

Table 2.2: Average RMS accuracies with standard deviations for the biplane fluoroscopic experiments (Section 2.5.1 and 2.5.3). The first row indicates the accuracy of the ground truth, second row gives the results of the 3D-2D rigid registration of the ground truth shape, third row shows accuracies achieved with the 3D-2D non-rigid registration of the SSM, and last row shows the best achievable representation of the 10 subjects using the SSM. We indicate point-to-surface distance (P2S) and edge to surface distance (ES) in mm.

Method	P2S	ES
ground truth	-	0.42 ± 0.05
3D-2D rigid registration	0.68 ± 0.28	0.46 ± 0.16
3D-2D non-rigid registration	1.68 ± 0.35	0.96 ± 0.39
3D-3D SSM to ground truth	0.78 ± 0.07	0.84 ± 0.22

2.5.4   Fluoroscopic sequence

Experiment   In the fourth experiment the applicability of the method to fluoroscopic sequences was investigated using a biplane fluoroscopy sequence of one subject (Section 2.3). Ground truth shape and pose were available for every fourth frame. Accordingly, the in-between frames were left out from the experiment. We performed both rigid registration of the ground truth bone as well as non-rigid fitting of the SSM 10 times on the sequence. In each frame, the algorithm was initialized with the shape of the previously fitted frame and the extrapolated pose calculated as the last pose updated with the last motion vector. We started fitting on the second available frame, where the initialization occurred with the mean shape and ground truth pose
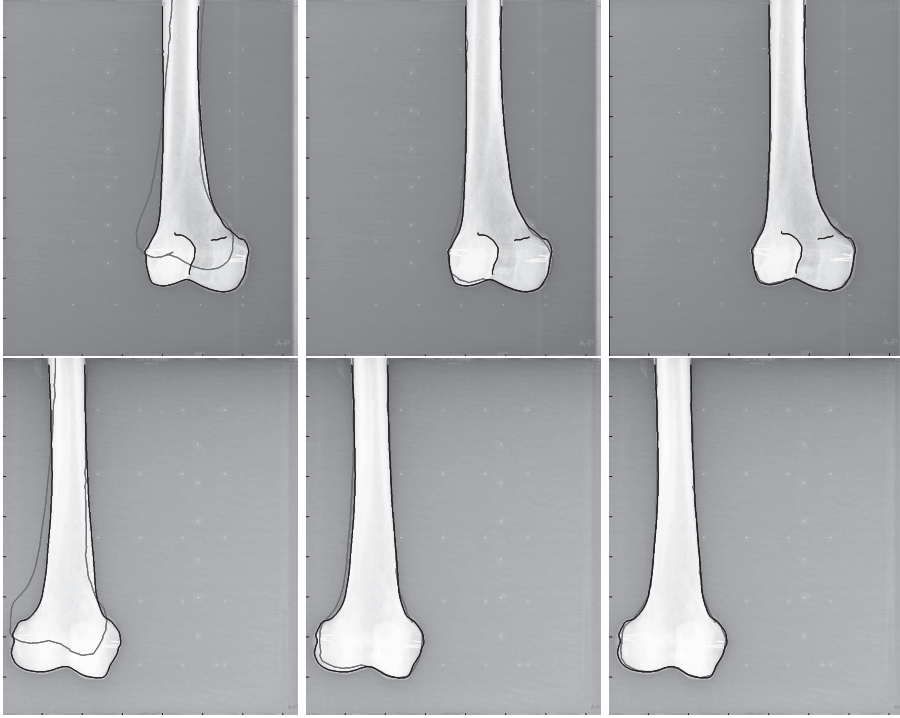
Figure 2.4 (color version on page 148): An example fit of Cadaver 1 in Anterior-Posterior pose. Canny edges are overlaid on the X-rays in black, red is the projection of the model. Top row: Left images, Bottom row: right images. From left to right: initialization $(P2S = 6.85mm)$, rigid fit with mean shape $(P2S = 2.3mm)$, non-rigid fit of the SSM $(P2S = 1.56mm)$.

of the first frame. In the SSM experiments we optimized pose (translation, rotation, scaling) as well as shape parameters in each frame.

Results Both rigid and non-rigid fitting accuracies and their standard deviation are given in Figure 2.7. The bar charts show the convergence rate calculated from the 10 trials. Higher errors occurred at the begin and the end of the sequence, due to the knee partially being outside the FOV, and additionally overlap of the two legs deteriorated the edge-map. Example frames from the graphs of Figure 2.7 are shown in Figure 2.8. Movie 2 of the supplemental material shows the sequence for a successful reconstruction[6].

---

[6]http://bigr.nl/movies/nbaka/movie2.gif
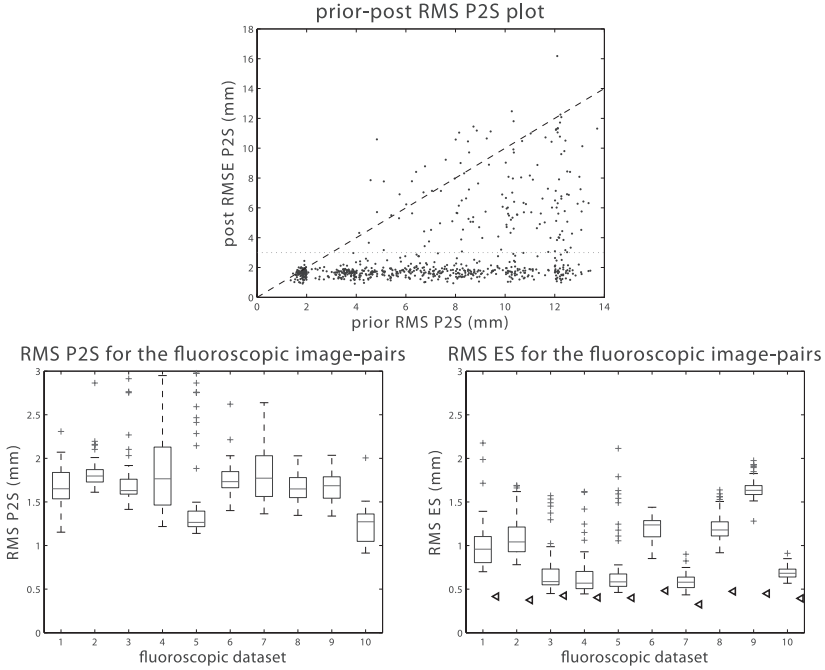
Figure 2.5: Non-rigid SSM fitting results on 10 fluoroscopic image pairs. Top: Convergence plot. Each dot represents a separate fitting with the initial P2S distance on the 'x' axis, and the post-registration distance on the 'y' axis. The red dotted line shows the convergence threshold, the dashed line is the no improvement line. Bottom left: Box-plot of P2S distances per fluoroscopic dataset for all converged cases, Bottom right: Box-plot of ES distances per fluoroscopic dataset for all converged cases. Black triangles mark the ES distance of the ground truth.

2.5.5 Influence of the Canny parameters

Experiment In the last experiment we assessed the influence of the choice of Canny parameters on the fitting accuracy and success rate. Besides the relevant object edges, patient scans contain edges due to noise, tissue overlap, and surrounding tissue. We therefore investigated the relation of the amount of object edges compared to all edges in the image (percentage of object edges, POE), and of the *amount* of object edges compared to the full silhouette (percentage of visible silhouette edges, PVSE) to the fitting result [7]. Edges were marked to belong to the object if they were within a region around the ground truth projected contour. The extent of this region varied with σ, accounting for the shift in the contour position and the ES distance of the ground truth. In some cases also noise edges fell into this region, generating a PVSE

---

[7]we do not study the edge *position* with respect to the smoothing parameter

Table 2.3: Mean and standard deviation of absolute pose distances of all converged reconstructions of the rigid registration with the ground truth CT shape (first experiment) and non-rigid fitting of the SSM (third experiment) on the 10 fluoroscopic datasets. Translation is in the world coordinate system [mm], rotations are in a local coordinate system around Lateral-Medial axis (LM), Anterio-Posterio axis (AP) and Proximal-Distal axis (PD) in degree units, and scaling errors are in percentage

Parameter	rigid CT bone	non-rigid SSM
x (mm)	0.29 ± 0.3	0.68 ± 0.6
y (mm)	0.26 ± 0.22	0.92 ± 0.71
z (mm)	0.45 ± 0.46	0.99 ± 0.91
Rot LM °	0.44 ± 0.45	0.93 ± 0.72
Rot AP °	0.38 ± 0.35	0.78 ± 0.68
Rot PD °	1.19 ± 1.01	2.19 ± 1.73
scale %	1.43 ± 0.55	1.55 ± 1.25

higher than 100. Furthermore, we investigated the effect of Canny parameters on the fitting accuracy by evaluating 12 different Canny parameter configurations.

Results The spread of the POE and PVSE for the different Canny parameters are illustrated in Figure 2.9. Colors code for Canny smoothing, while line style indicates the employed threshold value. A negative correlation between the measures can be observed: the more of the silhouette is visible, the smaller is the percentage of all edges in the image that is relevant. Results of the non-rigid fitting in Figure 2.9 indicate that a relatively high amount of relevant edges in the image results in a larger convergence radius. Also, the larger the portion of the silhouette present in the image, the better the reconstruction accuracy. The rise at the end of the distance curve on the right of Figure 2.9 shows deterioration of the fit due to noise. Table 2.4 gives an overview of the performance with varying canny parameters.

Table 2.4: RMS P2S accuracy and 90% convergence interval results of the SSM fitting to the 10 fluoroscopic biplane X-rays, using different canny parameters

sigma	3		4		6		8	
threshold	conv	acc.	conv	acc.	conv	acc.	conv	acc.
0.02	4.18	1.75	4.92	1.72	6.77	1.74	1.55	2.03
0.04	9	1.65	8.72	1.68	8.45	1.74	7.24	2.22
0.06	7.44	1.86	8.26	1.86	9.03	1.92	1.83	2.28

2.6 Discussion and Conclusions

We presented a method to reconstruct bony anatomy from two (or more) calibrated X-ray images utilizing a statistical shape model and a 3D distance based objective
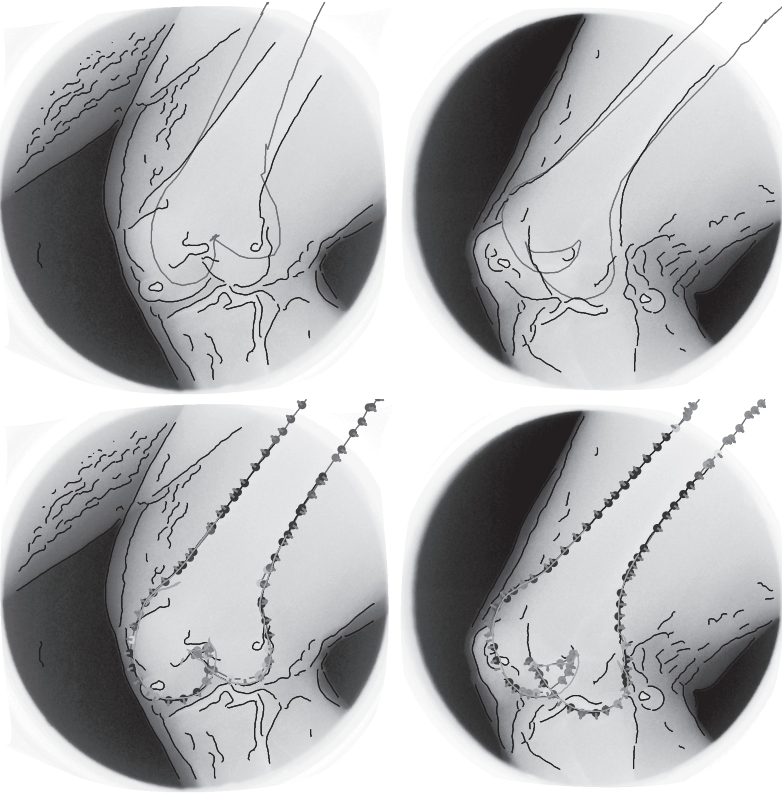
Figure 2.6 (color version on page 150): Top row: initialization Bottom row: matching scores (zero - dark blue; one - dark red) of the silhouette landmark points after SSM fitting. Ground truth shape is overlaid in green.

function using Canny edges in the projection images. Except for the initial pose estimate, the method performs reconstruction automatically.

We performed several experiments to investigate the performance of the pose and shape fitting of the distal femur from biplane images. The first two experiments represented simplified problems containing pure pose estimation of the ground truth shape, and pose and shape estimation in the ideal case of clear boundaries. In the third experiment we evaluated the accuracy of the overall proposed methodology using biplane fluoroscopic image pairs, which represent a realistic working environment for the method. Experiment four is a feasibility study showing the potential of the method for motion tracking from biplane fluoroscopy. Lastly, the influence of the Canny parameter settings on the reconstruction was evaluated.

The first experiment demonstrated that the rigid fitting of our method with the proposed data fit term obtained results comparable to those of the rigid fitting of ground truth pose (Figure 2.3, Table 2.2). The results also indicated that the rotation
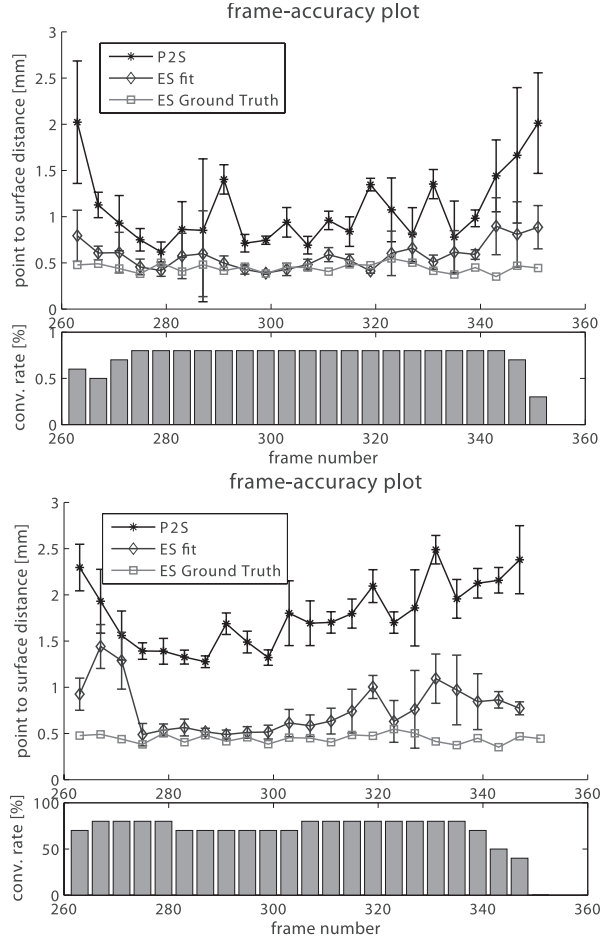
Figure 2.7: Results on a biplane fluoroscopic landing sequence. Fitting started at frame 263, and was performed on every 4th frame. The bar chart illustrates the convergence rate based on 10 trials, while the top graphs show mean, and standard deviation of the RMS distances of converged fits. Top: Rigid fitting results of the ground truth shape, Bottom: Non-rigid fitting results of the statistical shape model. A visual impression of a few frames of this fit is given in Figure 2.8.

around the longitudinal axis of the femur was the most difficult parameter to estimate (Table 2.3), which indeed has the smallest effect on the projected silhouette.

The non-rigid fitting accuracy of the method with an SSM was assessed in experiment two and three. The second experiment was signified by high contrast and a small angle between the images (35°). The third experiment was conducted on about 70° biplane fluoroscopic images, which due to the high-speed camera acquisition in-
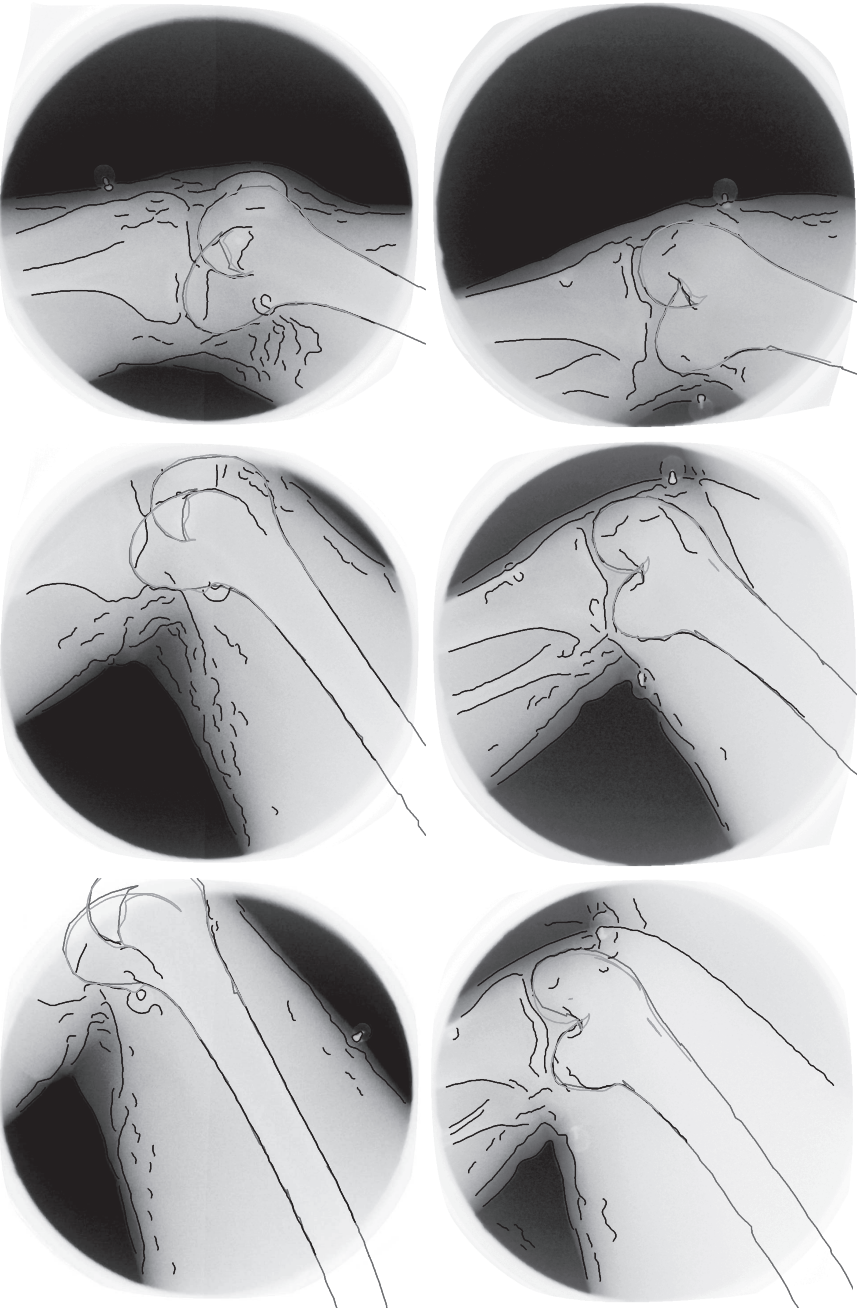
Figure 2.8 (color version on page 151): Examples of the SSM fit on the fluoroscopic soft landing sequence. Columns from left to right are frame 275, frame 307, and frame 331. Top row: Left images, Bottom row: Right images. Canny edges are overlaid on the X-ray with black. Red is the projection of the fitted shape, while green is the ground truth.
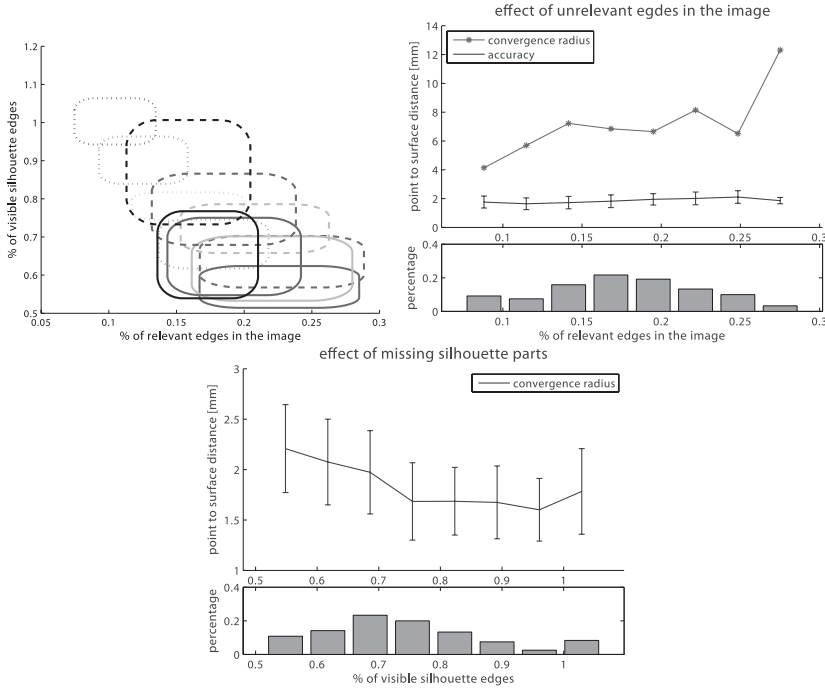
Figure 2.9 (color version on page 152): Effect of Canny parameters on the 10 fluoroscopic image-pairs. Top left: percentage of object edges, and percentage of visible silhouette edges for different Canny parameters. Colors code for the amount of Canny blurring: red $\sigma = 8$ pixels, green $\sigma = 6$ pixels, blue $\sigma = 4$ pixels, black $\sigma = 3$ pixels; and line style shows the threshold: 0.02 dotted, 0.04 dashed and 0.06 straight line. Top right: Convergence radius and accuracy of the 3D-2D non-rigid registration versus the percentage of object edges. Bottom: Convergence radius and accuracy of the 3D-2D non-rigid registration versus the percentage of visible silhouette edges. The histograms show the amount of test cases generating the plots.

hibit lower quality than standard fluoroscopy. The reconstruction of 10 subjects in the latter experiment resulted in an average RMS P2S accuracy of 1.68 mm (Table 2.2), and our experiments on two cadaver femora yielded an RMS P2S accuracy of 1.68 mm as well (Table 2.1).

These point-to-surface non-rigid fitting accuracies comprise several sources of errors. One important source is the model quality. The best possible representation (RMS P2S) of the ground truth shapes by the SSM was 0.78 mm and 1.0 mm for the fluoroscopic, and cadaver experiments respectively. Approximately half of the errors can thus be explained by this source. Further sources of error, also experienced in the rigid registration of the ground truth shape, are sparseness of the imaging setup; partial edges on the images; and presence of irrelevant edges in the edge map. Noise

edges may cause local minima affecting mainly the convergence radius. Partial edges on the other hand may cause false global minima in the pose as well as the shape parameter landscape. According to the rigid fitting results these sources are responsible for an RMS P2S error of approximately 0.68 mm for the fluoroscopic images (Table 2.2). Results also suggest an orientation dependency of the reconstruction accuracy (Table 2.1). This trend might be explained by the small angle (35°) between the biplane images causing various shape characteristics being invisible in certain scan orientations (e.g. the epicondyles in the lateral view).

In our experiments the cadaver reconstruction performed similar to the biplane fluoroscopy reconstruction. The smaller angle between the projections, and the more unusual shape of the cadaver bones due to their old age was most probably compensated by the larger FOV and the fact that cadaver X-rays contained the entire silhouette of the distal femur surrounded by almost no noise edges . The amount of visible silhouette edges was shown to have a positive influence on the non-rigid fitting accuracy (Figure 2.9). Often, the rigid fit of the mean shape outperformed the non-rigid fit to a partial silhouette in healthy young subjects. Nonetheless, non-rigid fitting led to significant improvements when the entire silhouette was visible, such as in the cadaver experiments.

The results of the proposed method for non-rigid shape reconstruction are comparable to previously published work on distal femur reconstruction from sparse projection images. Fleute et al. reported an accuracy of 0.99 mm RMS P2S with 90° biplane simulated silhouettes of the distal femur [44]. It is unclear from how many simulated cases this result was derived. Laporte et al. reported an RMS P2S accuracy of 1.4 mm on eight dry cadaver femora by elastic 3D model deformation to semi-automatically selected and region matched 2D contours [74]. Tang et al. reported a mean RMS of 1.72 mm and 1.95 mm on two cadaver femora with soft tissue [119]. They fitted a hybrid atlas to three projection images by a DRR matching scheme. We found no other study than ours evaluating non-rigid reconstruction of distal femora from in-vivo X-rays. While previous works require manual annotation of object contours [44, 74], and contour correspondence assignment [74], our method solely expects a sufficiently close initial pose estimate. Care should be taken, though, when directly comparing the quantitative numbers, as all papers employ different acquisition protocols, test subjects, and training datasets, FOVs, etc.

Pose estimation in the non-rigid SSM fitting experiments resulted in a good accuracy for translation, and in-plane rotation. Similar to the rigid registration of the ground truth shape, the most difficult pose parameter to infer from the biplane data was the rotation along the longitudinal axis of the femur (Table 2.3).

In the 3D reconstruction of fluoroscopic sequences we initialized the method with the shape and pose calculated from the previous frames, such that manual initialization was only provided in the first frame. Results using one sequence with an effective frame rate of 125 frames/second showed that the convergence radius of rigid and non-rigid reconstruction (Figure 2.3 and 2.5) was sufficient for motion reconstruction requiring one initial pose per sequence in 80% of the cases (Figure 2.7). Results also indicated a dependency of accuracy (and convergence rate) on the extent of the distal bone visible in the FOV (Figure 2.7). This experiment shows the potential

of the proposed method to be used with fluoroscopy, however, the current method was not optimized to sequence reconstruction as no prior knowledge about a constant shape throughout the sequence was used. A more sophisticated adaptation of the method to fluoroscopic sequences is planned as future work.

The presented method relies on the Canny edge map of the X-ray images. The effect of the choice of Canny parameters was assessed in the last experiment. In conclusion, a smoothing parameter of $\sigma = 3.0$ pixels and an upper threshold of $th = 0.04$ performed the best on the 10 fluoroscopic datasets used in this chapter (Table 2.4). We also conclude that there was no large difference in performance when applying slightly altered parameter values (e.g. $\sigma = 4.0$ pixels), ensuring that the parameter settings exhibit a high resemblance in the amount of noise edges, and object edges present in the edge-map. This is visualized as large overlaps in Figure 2.9.

By penalizing deviation from the mean shape in the optimization we limit the reconstruction method to healthy and mildly pathologic bones. Cases of severe fracture, and severe osteoarthritis cannot be reconstructed. However, most often knee replacement surgery is performed in an early stage of osteoarthritis, before large bone remodeling occurs. In this chapter both training, and test subjects lacked symptoms of bone pathology.

Improvements over the current method may be achieved by application of additional features besides the Euclidean distance and the gradient angle for point correspondence and/or weighting. Another extension of our work could be an active search of contours (such as dynamic programming) in the region around the shape estimate based on the Canny edges. This would provide a full silhouette, which might improve reconstruction accuracy. Furthermore, we currently model the entire femur bone, which is beneficial as information from the entire FOV can be used regardless of FOV size. However, in consequence variations of the shape in the non-imaged areas have to be modeled as well. An interesting extension could be fitting a FOV dependent shape model integrating benefits of both region-focused and FOV focused approaches.

The current approach has a computation time of about five minutes for a nonrigid 2D-3D reconstruction using the SSM on a 2.4 Ghz, 2GB laptop running an unoptimized Matlab implementation.

In conclusion, we presented a method, which is able to reconstruct 3D shape from two or more projection images with minimal user interaction. We tested the method on low quality biplane in-vivo fluoroscopy and achieved an accuracy of 1.68 mm (RMS P2S), which is comparable to methods in recent literature. Rigid pose estimation with the ground truth shape resulted in a sub-millimeter average RMS P2S accuracy (0.68 mm). The convergence radius of the method was shown to be sufficiently large to reconstruct shape and motion from a fluoroscopic sequence in 80% of the cases with minimal user interaction involving one manually initialized pose per sequence optimization.

2.7   Acknowledgments

Chapter 3

Statistical shape model based femur kinematics from biplane fluoroscopy

Abstract

Studying joint kinematics is of interest to improve prosthesis design and to characterize post-operative motion. State of the art techniques register bones segmented from prior CT or MR scans with X-ray fluoroscopic sequences. Elimination of the prior 3D acquisition could potentially lower costs and radiation dose.

Therefore, we propose to substitute the segmented bone surface with a statistical shape model based estimate. A dedicated dynamic reconstruction and tracking algorithm was developed estimating the shape based on all frames, and pose per frame. The algorithm minimizes the difference between the projected bone contour and image edges. To increase robustness, we employ a dynamic prior, image features, and prior knowledge about bone edge appearances. This enables tracking and reconstruction from a single initial pose per sequence.

We evaluated our method on the distal femur using eight biplane fluoroscopic drop-landing sequences. The proposed dynamic prior and features increased the convergence rate of the reconstruction from 71% to 91%, using a convergence limit of 3 mm. The achieved root-mean-square point-to-surface accuracy at the converged frames was 1.48±0.41 mm. The resulting tracking precision was 1-1.5 millimeter, with the largest errors occurring in the rotation around the femoral shaft (about 2.5° precision).

3.1 Introduction

Kinematic joint measurements are performed for various reasons, such as to relate joint anatomy to function [36, 122], to characterize pre- and post-operative joint motion [1], and to enable better prosthesis designs [47].

Mono- and biplane X-ray fluoroscopy have been extensively used in kinematic studies. Post-operative knee motion, for example, has been assessed by tracking the implants on fluoroscopic sequences [10, 47]. Such tracking methods utilize the shape of the implant as designed by the manufacturer. These methods have later also been applied for the tracking of natural joints. The required subject bones were segmented from 3D MR [1] or CT images [4, 36, 122], and subsequently registered per frame to the fluoroscopic sequence. Registration was either performed manually or by employing a 2D-3D rigid registration algorithm. A recent review on 2D-3D rigid registration algorithms can be found in [84]. Though sufficient accuracy can be achieved in this way, bone segmentation and per-frame registration may be laborious, and the prior CT acquisition exposes the subjects to additional ionizing radiation. Elimination of a prior CT or MRI imaging would therefore be beneficial, as it may lower radiation dose and acquisition costs.

Methods for bone shape reconstruction from a small number of X-ray images which do not require a 3D acquisition have been proposed in the literature. In these methods statistical shape models (SSMs) have been used to restrict the reconstruction to plausible shapes. The shape model is fitted to the X-ray images based on image intensity or other image features. Intensity based methods optimize intensity similarity between the X-ray images and digitally reconstructed radiographs (DRR) [60, 119, 134]. They may be sensitive to intensity changes in the fluoroscopic images (for example caused by overlap of the other leg in knee fluoroscopy) and can be computationally expensive. Feature based methods minimize the distance between the object edges in the image and the model silhouette [7, 12, 38, 44, 72, 73, 139]. Feature based methods either require a well defined object contour in the image or an initialization close to the desired solution. Both image and feature based 2D-3D reconstruction methods result in a pose and a shape estimate.

Application of a 2D-3D reconstruction algorithm for the kinematic tracking case was first proposed in [7]. However, per frame application of the above methods to derive kinematics is not optimal, as the bone shape is expected to stay constant. Chen et al. proposed to reconstruct the shape based on the first frame, and subsequently infer the kinematics using that shape estimate [24]. This solution to keep the shape constant is expected to be prone to the quality of the first frame. Furthermore, they derive kinematics by optimizing parameters of a motion model. While this approach may distinguish between normal and diseased subjects, it is less suited for studying the exact joint motion. Therefore, we propose estimating the shape based on all frames, and estimating pose per frame without the constraint of a motion model. This, however, is challenging due to the large motion between frames, requiring the automated method to be robust and have a large convergence range.

For the model fitting we propose a feature based 2D-3D reconstruction method, minimizing the distance of the projected model contour and the edges in the under-

lying image. To increase the convergence radius of the edge based reconstruction method we propose to assign a relevance measure to each edge-point based on a dynamic prior, image features, and prior knowledge about bone edge appearances. Enhanced edge information has already been used for point correspondence building by [48]. Our approach differs from theirs in that we employ image information combined with prior knowledge to derive an edge relevance rather than matching the corresponding properties of image edges and the projected bone silhouette. This enables incorporating multiple image features, such as intensity, edge length, etc. Yet, false correspondences are inevitable as certain parts of the object contour may be outside the field of view (FOV) or are not visible due to low contrast. To deal with these cases we improved the objective function presented in our earlier work [7] with an advanced outlier suppression mechanism.

The main contributions of this chapter are two-fold: 1) we propose a statistical shape model based tracking method that optimizes the shape based on the entire sequence and pose per frame, and 2) we propose to enhance edges with multiple image features incorporating prior information about edge appearances for establishing point correspondence. From a clinical point of view this results in a motion tracking method from biplane fluoroscopy that does not require the acquisition and segmentation of a prior CT or MRI. Furthermore, user interaction is minimized to determining one initial pose per sequence.

We evaluate our method on the distal femur using eight fluoroscopic drop-landing sequences by calculating the kinematic accuracy and point-to-surface fit accuracy. As a reference we also evaluate fitting accuracy using the segmented surface of a prior CT.

3.2   Method

The goal of the method is shape reconstruction and motion tracking from calibrated biplane fluoroscopic sequences. We use a SSM consisting of a mean shape and a set of parameters which change the shape in a plausible manner. These shape parameters for a given sequence and the pose for all time-frames are found by optimizing an objective function. The objective function consists of a data fit term encoding how well the projections of the model fit to the edges in the fluoroscopic images, and a shape term encoding the plausibility of the shape. The data fit term is calculated using corresponding edge-model point pairs determined on the image plane. The strategy used to optimize parameters of an entire sequence is described in Subsection 3.2.1. Further subsections give an overview of the statistical shape model, explain the proposed objective function, and describe the novel point correspondence determination.

3.2.1   Sequence optimization

To cope with sequences containing fast and large motion, we divided the optimization into two stages, as illustrated in Figure 3.1.

The first stage provides a robust initialization to the reconstruction. Here, the mean shape is rigidly aligned with the fluoroscopic frames resulting in a close to
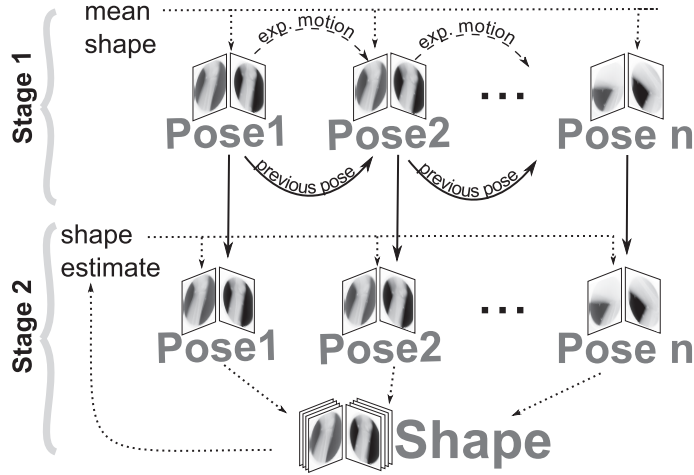
Figure 3.1: The two-stage sequence optimization. In the first stage the mean shape is rigidly aligned to the biplane frames. Here, expected motion is used in the correspondence building. The second stage alternates between pose and shape optimization. Dotted lines indicate fixed parameters, the dashed line indicates expected motion, bold lines indicate initialization, and the large gray text indicates optimization of the specified parameters.

optimal pose in every time frame. We fit the mean shape frame by frame with the previous frame's fitted pose as initialization of the next frame. For greater robustness edge enhancement contains a dynamic term. Further details on edge enhancement can be found in Section 3.2.4. In the first optimized frame we also approximate shape size, in further frames this scaling is kept constant.

The second stage comprises shape reconstruction and accurate tracking. In this stage both shape (including scaling) and pose parameters are optimized, resulting in one shape per sequence and one pose per time-frame. As the reference standard motion curves showed sudden changes in some cases, we assume in this stage no temporal dependency of the frames. A straightforward approach for this stage would be to optimize all pose and shape parameters at once. However, optimizing such large number of parameters poses a considerable computational burden, especially for long sequences. Instead, we propose to optimize pose and shape parameters in an alternating manner. Minimization of the objective function is thus first performed on the shape parameters on the entire sequence keeping the poses fixed, and then on the pose parameters of the separate frames keeping the shape fixed. Prior experiments showed that the estimated parameters converge in three or less alternations.

3.2.2   Statistical Shape Model

We build the SSM [27] by representing segmented training shapes with corresponding landmark points and their triangulation. Principal component analysis (PCA) of the procrustes aligned shapes is employed to derive the model $x \approx \bar{x} + \Phi b_x$, where x is the vector of the concatenated landmark coordinates of a shape, $\bar{x}$ is the mean shape, Φ contains the main modes of variation, and b_x is the shape parameter vector, i.e. the representation of the shape x in the parameter space.

3.2.3   Objective function

The proposed objective function E is an extension of the one presented in [7]. It quantifies how well the projection of the current shape estimate fits the edges of the underlying fluoroscopic data, and includes a penalty term to keep the recovered shape close to the training population. Accordingly, it is calculated as the sum of a data fit term E_{fit} and a prior term E_{prior} as illustrated in Fig. 3.2, and shown below:

$$E = E_{fit} + \lambda E_{prior} \quad , \tag{3.1}$$

where λ is a weighting parameter.

The Mahalanobis distance of the shape to the model mean is utilized as the shape prior penalizing improbable shapes:

$$E_{prior} = b_x^T \Sigma^{-1} b_x \quad , \tag{3.2}$$

where Σ represents the covariance matrix of the shapes in the parameter space.

As in [7], the data fit term E_{fit} is based on point correspondences between the projected shape silhouette and image edges, and acts in two ways: 1) it represents the goodness of fit based on the correspondences, and 2) it suppresses the effect of wrong correspondences. We used the perspective projection to project silhouette landmark points to the plane, with projection parameters known from prior calibration of the fluoroscopy system. Silhouette landmarks are shared by two triangles with one normal pointing toward and one pointing away from the X-ray source. Subsequently, the projected landmarks get a corresponding image edge pixel assigned as described in Section 3.2.4.

Similar to [7], we define a landmark point-edge pixel distance measure called matching score (Θ), that combines the Euclidean distance d between a projected landmark and its corresponding edge point with an angle distance w_α. Simplifying the matching score proposed in [7], in this work the Euclidean distance d is calculated on the plane.[1] The matching score is thereby calculated as

$$\Theta = 1 - w_\alpha \exp\left(-\frac{d^2}{\sigma_d^2}\right) \quad , \tag{3.3}$$

---

[1] Due to the orthogonal projection setup of our data, measuring the 2D distances rather than 3D is not expected to deteriorate results.
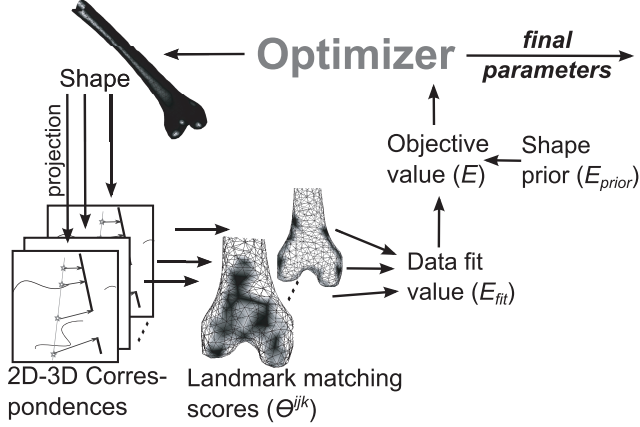
Figure 3.2: Parameter optimization. Based on correspondences between projected landmark points and image edges (Section 3.2.4), landmark matching scores are calculated per projection. These scores are combined to a data fit measure, which together with the shape prior forms the objective function. The optimizer iteratively determines the parameters such that the objective value is minimized. The number of fluoroscopic images used in the optimization depends on the optimized parameters as described in Section 3.2.1.

with

$$w_\alpha = \begin{cases} 0 & \text{, if } \alpha > 90° \\ \cos \alpha & \text{, otherwise} \end{cases} \tag{3.4}$$

where α is the angle between the projected normal vector of the 3D surface at a landmark point and the gradient of the fluoroscopic image at the corresponding edge location. The function Θ maps the Euclidean distances to the interval [0,1]. Due to its slow convergence to 1, it suppresses the effect of far correspondences on the optimization. This suppressing behavior is controlled by σ_d. The orientation term down-weighs correspondences with a large angle difference, and incorporates additional information about the landmark neighborhood.

Extending [7], we calculate the fit term as the weighted average of the landmark matching scores Θ^{ijk}, as

$$E_{fit} = \frac{\sum\limits_{i=1}^{n_{fr}} \sum\limits_{j=1}^{n_{pr}} \sum\limits_{k=1}^{n_k} W_f^{ijk} W_u^{ijk} W_e^{ijk} \Theta^{ijk}}{\sum\limits_{i=1}^{n_{fr}} \sum\limits_{j=1}^{n_{pr}} \sum\limits_{k=1}^{n_k} W_f^{ijk} W_u^{ijk} W_e^{ijk}} \quad , \tag{3.5}$$

where $k \in [1, n_k]$ represents the index of a projected landmark point inside the FOV of projection $j \in [1, n_{pr}]$ of frame $i \in [1, n_{fr}]$. The number of projections n_{pr} in a biplane system is two. The proposed new weights are as follows.

The weighting factor W_e is the edge relevance of the edge point corresponding to a landmark point. Edge relevance is further explained in Section 3.2.4, denoted as $\mathcal{F}(q)$.

The weighting factor W_u corrects for the non-uniform neighboring landmark distance in the projection plane. It is calculated as the average distance between a projected landmark point and its neighbors.

The weighting factor W_f suppresses the effect of landmark points close to the FOV border on the optimization. These landmark points are likely to be assigned wrong correspondences, as their correct correspondence might be outside the FOV. It is calculated as

$$W_f = 1 - exp\left(-\frac{D_{FOV}(p)^2}{\sigma_{FOV}^2}\right) \quad , \tag{3.6}$$

where $D_{FOV}(p)$ is the distance of the projected landmark point p to the FOV border, and σ_{FOV} is a parameter controlling the width of the region.

3.2.4 Correspondence building

A crucial part of the method is establishing the correct point correspondence between the silhouette landmark points of a shape (see Section 3.2.3) and the edge pixels of the fluoroscopic images. We applied Canny's method to derive the edges, which means hysteresis thresholding of the gradient magnitude image calculated by convolution with a Gaussian derivative kernel [21]. Numerous edges in the so derived Canny edge map are not associated with the object of interest and could potentially lead to false correspondences. Most irrelevant edges can, however, be suppressed for the correspondence determination by using prior knowledge. The corresponding edge point q of landmark point p can then be calculated as the one minimizing the cost function

$$\mathcal{C}(p,q) = D(p,q)^2 + w_G \, GD(p,q)^2 + w_F \, (1 - \mathcal{F}(q))^2 \quad , \tag{3.7}$$

where $D(p,q)$ is the distance between the projected landmark p and edge pixel q, $GD(p,q) \in [0,2]$ is the length of the difference vector of the projected and normalized surface normal of p and the normalized image gradient vector at q, and $\mathcal{F}(q) \in [0,1]$ is a measure of relevance for edge pixel q with 1 being highly relevant. See next paragraph for a detailed description. The contributions of the subparts of the function are balanced by weighting parameters w_G and w_F. As $C(p,q)$ is a sum of squares measure, the minimizing edge pixel q for a given projected landmark p can be efficiently found using spatial sub-division strategies, such as a KD-tree. The tree has to be built once per frame prior to optimization.

The edge relevance function $\mathcal{F}(q)$ is based on image features and motion dynamics. The image features are image measurements transformed to the interval $[0,1]$ using a sigmoid function

$$\mathcal{S}(x) = \frac{1}{1 + e^{\rho x + \nu}} \quad , \tag{3.8}$$

where x represents the measurement. The steepness ρ and offset parameters ν of the sigmoid are chosen for every feature separately, such that a high feature value reflects
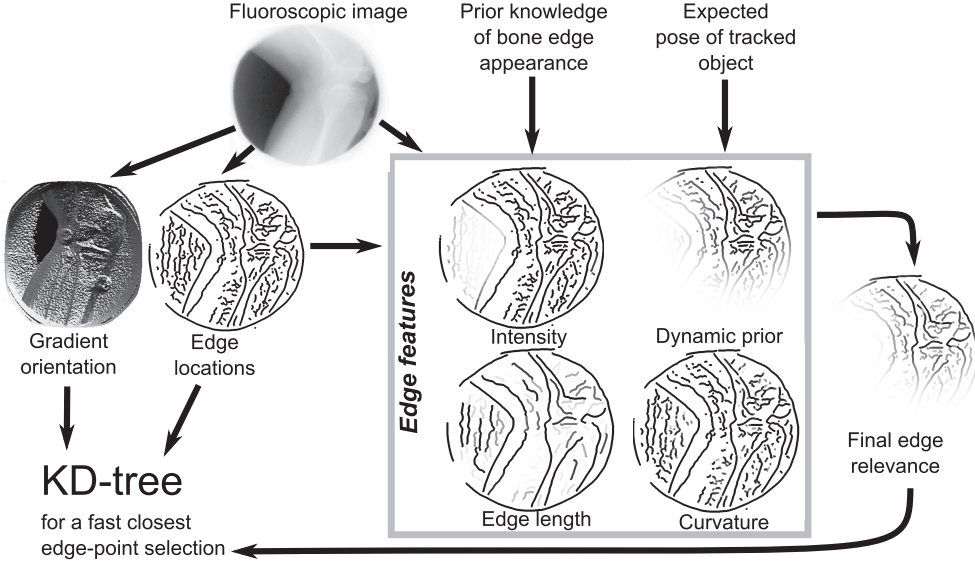
Figure 3.3 (color version on page 154): Correspondence building. Edge location, gradient orientation, and edge relevance are used in the correspondence determination. Edge relevance is a combination of edge features calculated from the original image, the Canny edge map and prior knowledge as explained in Section 3.2.4. Example images of one projection are shown with edges dilated for better visibility. The images are plotted with a linear colormap (1:black, 0:white).

high expected relevance. Finally, all features are combined to one relevance measure as described at the end of this section. An illustration of the correspondence building and the separate edge features is given in Figure 3.3. The following features and prior knowledge are proposed:

1. The intensity feature $\mathcal{F}_I(q)$ incorporates the prior knowledge that the bone lies inside the leg. Therefore, it results in higher relevance value for high intensities, implementing soft thresholding of the leg. The intensity feature is calculated via Eq. 3.8 with x being the image intensity at edge pixel positions.

2. The edge length feature $\mathcal{F}_L(q)$ encodes the prior knowledge that noise edges tend to be shorter than edges of real structures, such as bones, skin, etc. All pixels of an edge are assigned the same feature value calculated from Eq. 3.8 with x being the length of the containing edge.

3. The curvature feature $\mathcal{F}_C(q)$ encodes the prior knowledge that bone edges have low curvature. It is calculated from Eq. 3.8 with x being the smoothed image intensity based curvature sampled at edge locations.

4. Dynamics prior. Due to the high frequency acquisition of the fluoroscopic sequence, first order extrapolation of the pose (constant motion) provides a good approximation of the next pose. Therefore, edges are assigned the dynamics based edge relevance

$$\mathcal{F}_D(q) = \exp\left(-\frac{D(q)^2}{\rho_D^2}\right) \quad , \tag{3.9}$$

where $D(q)$ is the distance of the edge pixel q from the projected contour of the approximated next shape, and ρ_D determines the range of influence of the prior.

Combined edge relevance The final edge relevance is a combination of all features as shown:

$$\mathcal{F}(q) = \mathcal{F}_I(q)\mathcal{F}_D(q)\frac{\mathcal{F}_L(q) + \mathcal{F}_C(q)}{2} \quad . \tag{3.10}$$

Intensity and dynamics are used multiplicatively, greatly decreasing the likelihood of edges with low feature values. The curvature and the edge length features are combined by averaging. This decreases the effect of noise in the separate features.

The dynamic prior is used to increase the convergence radius in the first stage of the optimization (see Section 3.2.1). When such prior is not applicable, such as in the first two frames of the rigid fitting, or not desired, such as in the final non-rigid fitting, $\mathcal{F}_D(q)$ is set to one for all q.

The edge relevances of an image have to be computed once, and remain unchanged throughout the SSM fitting. Details of the feature parameter settings are given in Section 3.4.1.

3.3   Data

3.3.1   Femur shape model

The femur shape model was constructed from 43 femur segmentations of CT images. A detailed description of the data is available in [7]. Point correspondence across the training shapes was calculated using the GAMES algorithm [42] resulting in a landmarked mesh with an average landmark distance of 9 mm. We modeled the distal 70% of the femur leaving the femoral head out. This way we preserved all information that might be in the FOV while minimizing the number of modeled points (512 landmarks). Retaining 95% variance at model construction resulted in 32 out of 42 modes of variation.

3.3.2   Fluoroscopic data

We used a subset of the data of an ongoing study designed to give insight in 3D knee kinematics during jump-landing [122][2]. Subjects were asked to jump from a

---

[2]A different subset of the same study was used in [7]

40 cm high box to the ground in the field of view of a custom built biplane fluoroscopy system. The system was focused on the dominant landing support leg, and was constructed of two modified C-arms (Philips Medical systems, Best, NL) placed at an angle between 58-82 degrees, equipped with synchronized high-speed cameras recording at 500 frames/second with 8 bit bit-depth. Matching current practice, only every fourth time-frame was used for tracking creating an effective frame rate of 125 frames/second. Biplane fluoroscopic jump sequences were acquired for 10 healthy subjects with an image resolution of 1024x1024. Two sequences had one plane with resolution 800x600. The calibration procedure is described in [64] (method 4). We note, that because of the high-speed stereo acquisition and the 8 bit bit-depth of the cameras the fluoroscopic data used in this chapter is of lower quality than standard fluoroscopy.

3.3.3 Reference standard for the fluoroscopic data

All subjects underwent a CT scan with a FOV ranging from approximately 12 cm above to 12 cm below the joint line of the knee. Triangulated CT segmentations (extracted with Mimics, Materialize Inc., Plymouth, MI) were used as ground truth shapes. The reference standard pose (translation and rotation) was derived by 2D-3D registration of the ground truth shape to manually selected canny edges of the fluoroscopic images, using commercially available software for Model-Based Roentgen Stereophotogrammetric Analysis (MBRSA) (Medis-specials, Leiden, The Netherlands)(for details see [65]). The reference standard was created on every 4th time-frame.

None of the subjects of the fluoroscopic data was included in the training data of the femur shape model. A 3D-3D fitting of the SSM to the ground truth shapes revealed its intrinsic representation accuracy of 0.8 mm root mean square (RMS) point-to-surface distance.

3.3.4 Anatomic coordinate system

For evaluation we adapted the anatomic coordinate system proposed by [94]. This coordinate system approximates the mechanical axes of the femur for small FOVs (with missing femoral head). The medial-lateral (ML) axis was determined by fitting a cylinder to the posterior condiles. The anterior-posterior (AP) axis was calculated by taking the cross product of the ML axis and the mid-line of the femoral shaft. This mid-line was determined by the smallest inertial axis for the SSM, and by fitting a cylinder to the shaft points for the fluoroscopic ground truth shapes, as there the shaft was not long enough to consistently orientate the inertial axes. Finally the proximal-distal (PD) axis was set orthogonal to both the AP and ML directions. The origin of the coordinate system was defined at the mid-point of the cylinder on the ML axis.

3.4   Implementation details

3.4.1   Parameter settings

The proposed method requires four groups of parameters to be set prior to application. Parameters were set based on two sequences, and were subsequently used in all sequences and experiments. Results on these two training sequences are reported separately.

The Canny edges of the X-rays were computed with blurring $\sigma = 4$ pixels (matching the protocol that was used in obtaining the reference standard pose). The higher threshold of the gradient magnitude was set to 1.5 % of the maximum gradient magnitude of the image, and the lower threshold was set to 0.3 %. These settings resulted in acceptable femoral bone edges, but also many irrelevant and noise edges. These parameters were manually set with a large margin to ensure applicability for all sequences.

The offset and steepness parameters of the sigmoid functions of Section 3.2.4 were set based on the image properties. The intensity feature parameters were set such that the upper bound of intensity values outside the leg (30) and a typical soft tissue intensity in the leg (120) resulted in feature values of 0.05 and 0.99 respectively. The edge length feature parameters were set such that an edge of length 20 pixels (about 6 mm) and 80 pixels (about 2.5 cm) gave feature values of 0.05 and 0.95 respectively. The curvature feature is typically noisy due to the second image derivative involved in its calculation. Therefore, we set the parameters conservatively such that a curvature of 0 and 1 gave features values of 0.99 and 0.05 respectively. This filters mainly high curvature values out. We measure curvature in pixels, thus a curvature of 1 means a touching circle with radius $\sim$0.3 mm. The weights w_G and w_F were chosen as 200 and 100 respectively. This choice is a balance between distance, angle and edge relevance.

The width parameter σ_d of the fitting term in Equation 3.3 is gradually decreased during the optimization. While distant correspondences in the beginning may be due to a far initialization, they are likely to be outliers at the end. Therefore we set $\sigma_d^2 = 500$ mm^2 for translation optimization and $\sigma_d^2 = 200$ mm^2 for the translation and rotation optimizations. When fitting the ground truth shape σ_d decreases to $\sigma_d^2 = 30$ mm^2. The FOV mask width in Equation 3.6 was set to $\sigma_{FOV} = 30$ pixels.

The objective function has a factor λ, which determines the relative weight of the prior term. A high value of λ results in a rigid model, which is able to fit the image edges less accurately, while low value of λ may result in implausible shapes. We found $\lambda = 0.01$ a good compromise.

3.4.2   Numerical optimization

The optimization routine employed to minimize the objective function in both stages was a non-linear least squares solver as implemented in the optimization toolbox of Matlab (R2011b). We used a subspace trust-region method based on the interior-reflective Newton method. The iterative minimization was stopped when the step size decreased under a threshold, or the number of iterations exceeded a preset maximum.

In the first stage we employ a three-step optimization strategy enlarging the set of optimized parameters from translation only to translation and in-plane rotations, to finally including all 6 degrees of freedom.

3.4.3 Model up-sampling

As the data fit term of the objective function is only evaluated at landmark point positions, a coarse model might introduce local minima in the minimization. These local minima can be avoided by up-sampling the model. We employed in the experiments a 7 fold up-sampling, meaning that we sampled six extra points equally distributed on every model edge. The surface normal of the adjacent landmark points was interpolated to derive the normal for these new points. These additional points were then handled as landmark points throughout the matching.

3.5 Experimental setup

3.5.1 Evaluation scheme

Evaluation measures

The aim of our work is shape reconstruction and tracking. Accordingly we introduced two evaluation measures.

The point-to-surface (PS) distance between the estimated shape with the estimated pose and the ground truth shape in the reference pose evaluates the fit accuracy. PS errors were only calculated at the landmark points of the knee, limited by the FOV of the prior CT and the ground truth segmentation. The PS distance of a shape is given as the root mean squared (RMS) distances of its landmark points. Matching convergence was defined conform to [7] as an RMS PS error smaller than 3 mm.

The motion difference with respect to the reference standard indicates the tracking performance. Motion is defined relative to the starting frame. We report the mean (bias) and standard deviation (precision) of the motion differences between reference standard and reconstructed sequence per parameter, i.e. translation, rotation. Translation parameters show the translation of the origin of the femoral anatomical coordinate system, and rotation the separate degrees of rotation along the anatomical axes.

Frame categorization

Due to over-projection of the second leg and limited FOV, a varying amount of femur information is visible throughout a sequence. We assigned a category to every time-frame based on the information contained in the two projection images. Such categorization will allow us to study the relationship between tracking accuracy, and bone information contained in the images. Two independent observers rated the fluoroscopic images in a random order as either: 1. all necessary information for reconstruction of the position and orientation of the distal femur is in the FOV and

is visible, 2. pose of the femur is reconstructible despite large parts of the distal
femur not being visible, and 3. pose of the femur is hardly reconstructible, because
important parts of the distal femur are not visible. Every image was assigned the
average mark of the two observers rounded downward. Finally, a time-frame was
assigned to category 3 if at least one of its projection images was marked 3, 2 if both
projections were marked 2, and 1 otherwise.

3.5.2 Experiments

We performed three sets of experiments to evaluate different aspects of the proposed
method. The first experiment shows the fit accuracy and tracking performance of
the entire proposed method using the statistical shape model. The second set of
experiments gives an impression of the effect of the proposed sequence based shape
optimization and edge relevance terms on the fitting accuracy. In these experiments
parts of the proposed framework were left out, while leaving the rest of the method
unchanged, resulting in a separate evaluation of the different contributions of the
chapter. The third experiment shows the tracking performance in the simpler problem
when the segmented bone surface from CT is available. This case serves as the best
possible accuracy reference for the method without shape estimation.

For all experiments tracking was performed on all 10 available drop-landing se-
quences with 10 starting poses. We simulated the manual initialization by adding a
random displacement to the reference standard pose of the starting frame. The ran-
dom displacement was sampled from a uniform distribution of the interval $[-10, 10]$
per parameter, with metrics mm for translation and degree for rotation. Because
motion dynamics in the starting frame is not available to guide the fitting, selecting a
good quality frame is advantageous. The starting frame for all subjects was chosen as
the seventh frame after the femur enters the FOV. Due to the jumping trajectory this
frame is highly likely to contain the joint, and also part of the femoral shaft. Tracking
was then performed in both directions from the starting frame. We report results of
the eight test sequences and the two sequences used to set the method parameters
separately.

3.6 Results

Experiment 1

The results of the first experiment are summarized in Table 3.1, containing only
the eight sequences that were not used in the parameter optimization. We report
convergence rate, and mean accuracy with standard deviation of the time-points in
which the fitting converged. Additionally, Figure 3.7a depicts all fittings including
non-converged cases to get an impression of the spread of the accuracies over time.
Figure 3.5 shows the PS distances per subject for time-points in category 1 and
2. We added to this graph the training subjects (subject 1 and 2) as well to show
their performance compared to the test sequences. Figure 3.6 depicts two examples
of a fitting result, and a full sequence reconstruction of subject 7 is shown in the
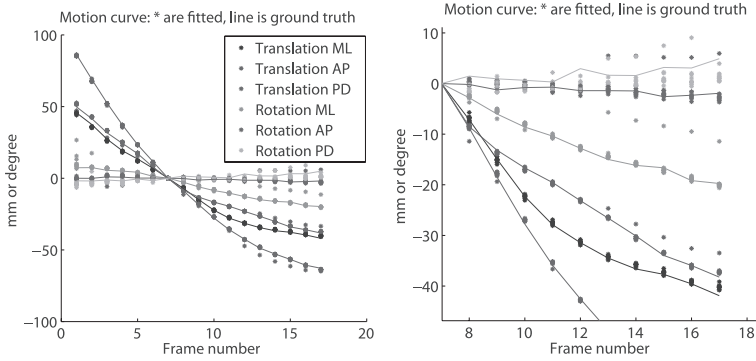
Figure 3.4 (color version on page 153): Detailed results of subject 7. Left: Motion curve with respect to frame 7. The solid line shows the ground truth motion, and stars represent the SSM fitting results of 10 runs with different starting poses. Right: a zoom-in to the time-points 8-17.

complementary movie (movie 3)[3]. Figure 3.4 shows the motion curve of subject 7 with respect to the starting frame. The solid line represents the reference standard motion, and stars represent the SSM fitting results of 10 runs with different starting poses. The wide spread of the longitudinal rotation (the rotation around the proximal-distal axis) illustrates the difficulty of optimizing that parameter. In our experiments an initialization within the radius of 8.4 mm RMS PS distance from the reference standard shape in the starting frame lead to a 90 % convergence rate. A manual initialization with such quality can be easily achieved [7].

Experiment 2

The convergence rate and PS accuracy of all frames for four methods are shown in Table 3.2. The first row indicates the performance of a method that optimizes the shape parameters on the entire sequence, but does not include the image features nor the dynamic prior. The second row shows the results when the dynamic prior was included in the fitting. The third row represents a method that includes edge relevance as proposed, but optimizes the shape parameters solely on the starting frame (frame 7). Finally, the last row represents the proposed method.

To get an impression of the overall results without the distinction of convergence, results of the first and last rows of the table are shown in Figure 3.7.

Experiment 3

Quantitative results of the rigid fitting of the ground truth shape are shown in Table 3.3. This table contains only sequences that were not used in the parameter optimization. Figure 3.8 shows the PS distances per subject for time-points in category

---

[3]http://bigr.nl/movies/nbaka/movie3.gif

Table 3.1: Results of the tracking with the SSM based on all fits of the eight test sequences. Motion and PS errors are only reported for the converged cases. Motion errors show the mean error (bias) with its standard deviation (precision). Frame categories are described in Section 3.5.2. Values are reported in mm and degree units in the anatomic coordinate system.

	Cat 1	Cat 2	Cat 3	All
nr	1260	90	150	1500
conv %	97	87	51	91
PS	1.43 ± 0.37	1.86 ± 0.53	1.88 ± 0.37	1.48 ± 0.41
Tr ML	0.16 ± 1.33	0.29 ± 3.34	0.39 ± 1.66	0.18 ± 1.53
Tr AP	0.03 ± 0.91	-0.26 ± 1.15	-0.62 ± 0.86	-0.02 ± 0.93
Tr PD	-0.43 ± 1.05	0.07 ± 2.00	-0.49 ± 1.66	-0.41 ± 1.17
rot ML	-0.05 ± 1.11	1.11 ± 3.66	0.44 ± 1.90	0.04 ± 1.46
rot AP	0.02 ± 1.02	-0.39 ± 1.78	0.84 ± 2.12	0.04 ± 1.18
rot PD	-0.86 ± 2.17	-2.79 ± 3.00	-1.98 ± 4.97	-1.03 ± 2.51
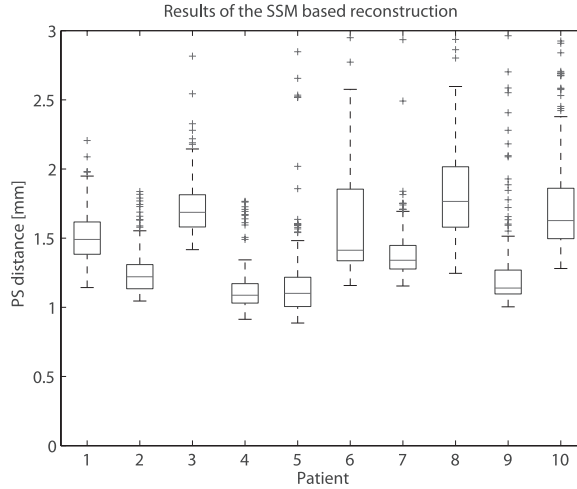
Figure 3.5: Boxplot of all converged frames of category 1 and 2 for the SSM fitting. Subjects 1 and 2 were used for parameter estimation.
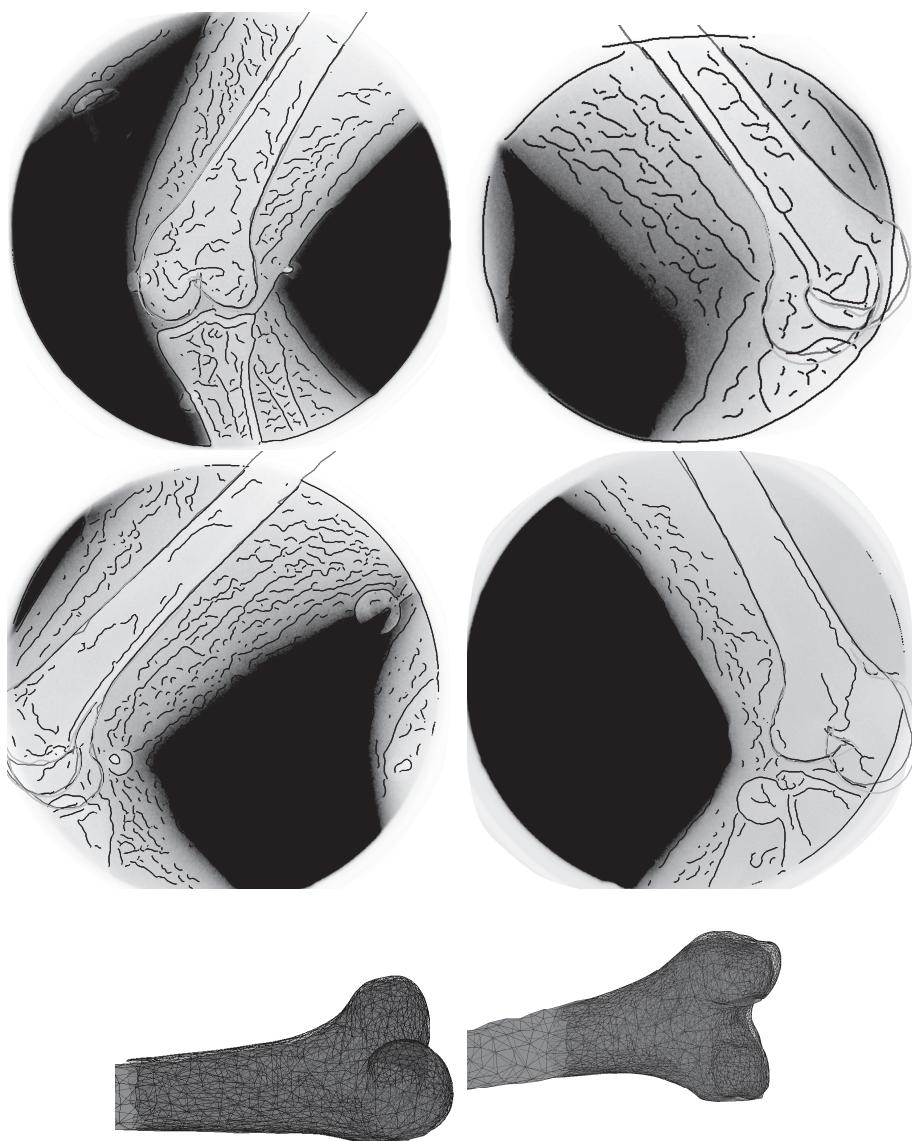
Figure 3.6 (color version on page 155): Two category 1 time-points of different subjects with overlayed contours of the fitted SSM (red) and the reference standard (green). From top to bottom: first plane, second plane, 3D surfaces of the reference standard (light blue) and the fitted SSM (red), rotated to see the differences. Left column shows a time-point with RMS PS distance of 1.20 mm (subject 7), and the right column shows a time-point with RMS PS distance of 1.74 mm (subject 3).
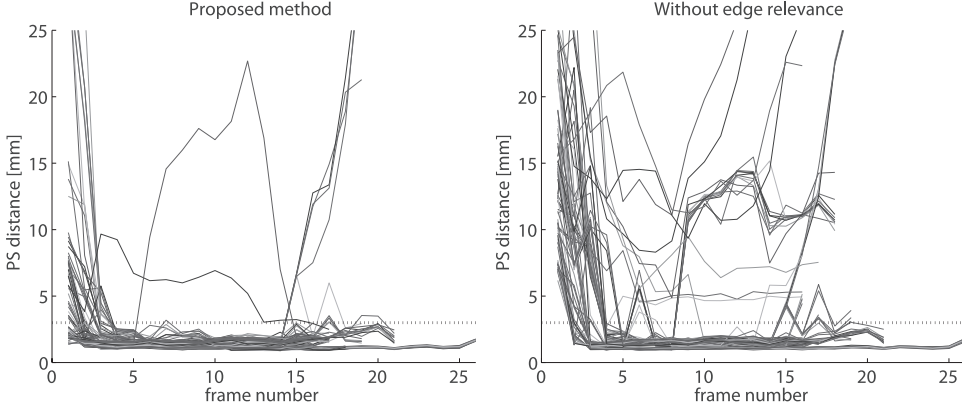
Figure 3.7: PS distances of all fitting tries of all test sequences plotted as a time-curve. The black dotted line represents the convergence limit. Left: proposed method, Right: tracking without the proposed edge relevance terms.

Table 3.2: Convergence rate and accuracy comparison of methods containing parts of the proposed method.

Shape opt.	Image features	Dynamic prior	conv. rate	PS distance
sequence	-	-	71	1.42 ± 0.36
sequence	-	+	82	1.44 ± 0.39
one frame	+	+	88	1.53 ± 0.46
sequence	+	+	91	1.48 ± 0.41

1 and 2. Here, the training sequences (subject 1 and 2) are also listed to show the difference in performance. Figure 3.9 shows two examples of a fitting result.

3.7 Discussion and Conclusions

We introduced an automated method for tracking the femur in biplane fluoroscopic sequences, that does not require the acquisition of a prior 3D image such as a CT scan. To validate the proposed method we report tracking and reconstruction results on eight drop-landing sequences. We ran our algorithm 10 times on all sequences with considerably different starting poses to demonstrate reproducibility in a wide range of initial poses. Besides evaluating the SSM based tracking, we also computed the performance of the tracking with the patient specific distal femur shape. The resulting accuracies serve as a reference for the case a 3D scan of the subject is available. Furthermore, we showed the usefulness of the sequence based SSM reconstruction and the edge relevance terms by assessing the performance of the method without these contributions.

Figure 3.8: Boxplot of all converged frames of category 1 and 2 for the rigid fitting of the ground truth shape. Subjects 1 and 2 were used for parameter estimation.
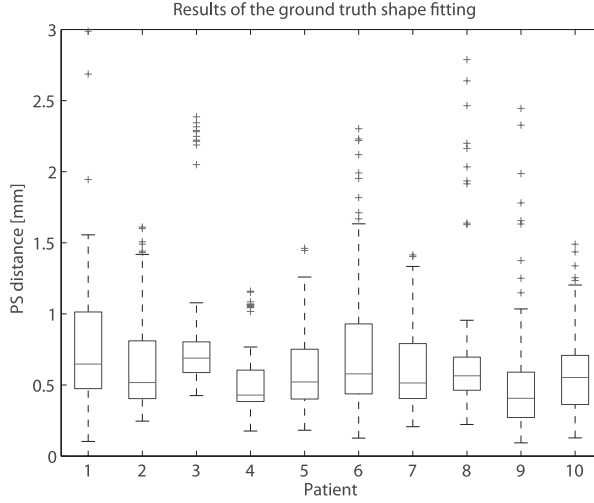
Table 3.3: Results of the tracking with the ground truth shape based on all fits of the eight test sequences. Motion and PS errors are only reported for the converged cases. Motion errors show the mean error (bias) with its standard deviation (precision). Frame categories are described in Section 3.5.2. Values are reported in mm and degrees in the anatomic coordinate system.

	Cat 1	Cat 2	Cat 3	All
nr	1260	90	150	1500
conv %	98	90	63	94
PS	0.57 ± 0.27	1.24 ± 0.76	1.03 ± 0.53	0.64 ± 0.39
Tr ML	-0.05 ± 0.62	-0.04 ± 0.88	-0.14 ± 0.58	-0.06 ± 0.63
Tr AP	0.06 ± 0.52	0.48 ± 0.73	-0.23 ± 0.93	0.06 ± 0.58
Tr PD	-0.12 ± 0.43	0.20 ± 1.42	0.41 ± 0.81	-0.06 ± 0.58
rot ML	0.09 ± 0.70	0.15 ± 1.56	0.93 ± 1.33	0.15 ± 0.86
rot AP	0.00 ± 0.63	0.33 ± 0.88	-0.19 ± 1.00	0.01 ± 0.69
rot PD	-0.34 ± 1.07	-0.17 ± 2.33	-0.13 ± 1.74	-0.32 ± 1.23

The proposed SSM based femur tracking converged in 91% of all frame optimizations (Table 3.1). Convergence is an important measure for assessing the capabilities of an algorithm for automated tracking. Breaking the results down into categories, we conclude, that the smaller amount of bone is visible in a frame the less the convergence rate becomes. This is especially visible at the first few frames of the sequences, when the femur is largely outside the FOV (Fig.3.7). For best category frames (which constitute 84% of all frames) convergence rate was 97%, which we consider high enough
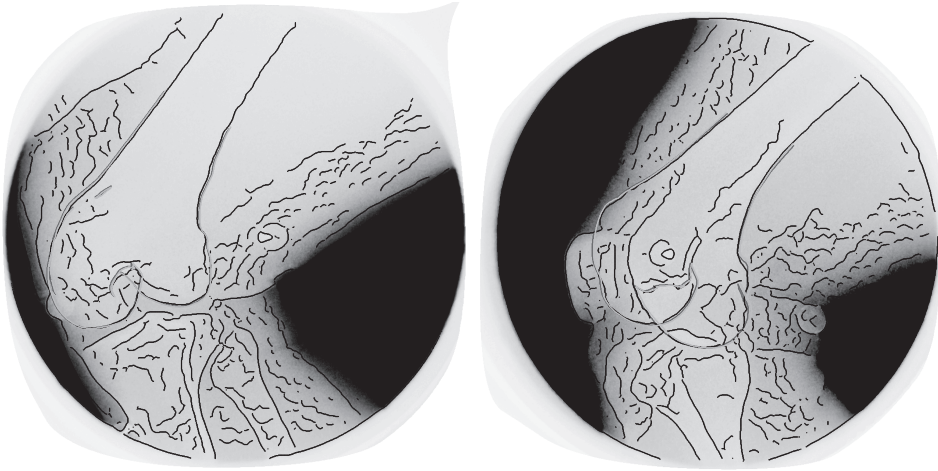
Figure 3.9 (color version on page 153): First and second planes of a category 1 time-point of subject 4 with the Canny edges and the fitted ground truth shape being overlaid. Red is the fitted ground truth shape, and green is ground truth shape in the reference standard pose. This time-point has PS distance of 0.65 mm.

for practical use. The comparison experiments showed furthermore, that the incorporation of the proposed dynamic prior and the image features increased the overall convergence from 71% without the proposed features to 91% with (Table 3.2). The increase was consistent for all frame categories (Fig. 3.7). The shape optimization based on the entire sequence as proposed in this work not only increased the convergence radius of the reconstruction compared to the single frame reconstruction approach, but also resulted in a higher accuracy (Table 3.2).

The SSM fitting resulted in an average RMS PS accuracy of 1.48 mm on all frames (Table 3.1). The categorical separation of frames indicates a better reconstruction performance of frames containing larger visible parts of the femur. The accuracy achieved at the best category frames was 1.43 mm. Part of the error is due to the small training set size of the statistical model, as can be seen from the model representation accuracy of 0.8 mm RMS PS (Section 3.3). This represents the best achievable accuracy with our model.

In the tracking the largest pose errors were made in the longitudinal rotation (around the posterior-distal axis, Table 3.1 and Fig. 3.4). Accuracy of this parameter increased by more than a factor of two when applying the ground truth shape rather than the SSM. The bone surface, including the condyles, have a relatively short distance to the posterior-distal axis. Therefore, rotation around this axis causes only small changes in the contours. In addition, the statistical shape model allows variations in bone outline, which might compensate for such small contour changes due to rotation. Further differences in the models that could influence the rotation accuracy are the coarser resolution of the shape model, and its longer shaft compared

to the ground truth shapes. Longitudinal rotation changes the appearance of the condyles relatively more than that of the shaft. A longer shaft therefore might down-weight the effect of rotation on the fit measure.

The RMS PS distance of 1.43 ± 0.37 mm for reconstruction of high category time-frames is comparable to single time-point biplane distal femur reconstruction results found in the literature, despite the lower quality fluoroscopic images used in this study. [44] reported an accuracy of 0.99 mm RMS PS with 90° biplane simulated silhouettes of the distal femur. It is unclear from how many simulated cases this result was derived. [74] reported an RMS PS accuracy of 1.4 mm on eight dry cadaver femora by elastic 3D model deformation to semi-automatically selected and region matched 2D contours. [119] reported a mean RMS of 1.72 mm and 1.95 mm on two cadaver femora with soft tissue. They fitted a hybrid atlas to three projection images by a DRR matching scheme. [48] report a mean PS distance of 0.86 mm for the reconstruction of the entire femur from two orthogonal projections of three pairs of cadaver femora using free-form deformation on a generic model. [7] reported a mean RMS PS distance of 1.68 ± 0.35 mm on category 1 in-vivo biplane fluoroscopic single time-point reconstructions of 10 subjects. While [44] and [74] require manual annotation of object contours, and [74] requires a contour correspondence assignment, [7] and the presented method solely expect a sufficiently close initial pose estimate. Care should be taken, though, when directly comparing the quantitative numbers, as all papers employ different acquisition protocols, test subjects, training datasets, FOVs, etc.

The results of the method with the SSM were also compared to the situation in which the ground truth shape is known. In this case the matching had sub-millimeter overall PS accuracy (Table 3.3). Fitting to low quality frames, such as the first and second frames of the sequence often did not converge. In some medium and low quality frames we observed furthermore, that though the available information was fitted well, the ground truth and our result differed outside the FOV.

The parameters of the image features used during tracking were set based on two sequences randomly chosen from the ten available ones. Fig. 3.5 and Fig. 3.8 indicate that the method performance on these two sequences was similar to the test sequences. Therefore, we believe that using other sequences for training would result in similar performance.

In the current study, only every fourth frame of the sequences was used. We chose this to match current research practice, where such a frame rate was found sufficiently large to characterize the motion. Also, fluoroscopic acquisition at the original frame rate is very rare. By temporal down-sampling we demonstrate the applicability of our method to a wider range of fluoroscopic setups, operating at considerably lower frame rates.

The reference standard motion curves of one subject contained abrupt changes at one time-point. It is yet unclear if these abrupt changes are due to knee motion, image artifacts, or inaccuracies in the reference standard. Therefore, a smoothness constraint was to this end not included in the fitting.

Reconstruction of an average length sequence on a current laptop computer took about 6 hours in our un-optimized matlab based implementation. The most time-

consuming part of the method is the shape optimization on all frames. As this part is highly parallelizable, computation time can be greatly decreased if necessary.

There are several limitations to the current study. Bone implanted marker based Roentgen Stereophotogrammetric Analysis (RSA) type gold standard is hard to acquire for in-vivo measurements. Therefore, we compared the results of the proposed method with a reference standard pose obtained by a manual annotation based 2D-3D rigid registration algorithm rather than a real gold standard. Though this algorithm has been reported to be highly accurate in [65], evaluation of its performance in the current setup has not yet been performed.

Furthermore, consistent with the aim of motion tracking, we evaluated the tracking performance calculating the relative pose. This measure is dependent on the starting frame, as the errors of the starting frame are propagated throughout the sequence, hence could lead to different results with varying starting frames. This is not the case with the PS distance.

Ultimately, we would like to track and evaluate the relative motion between the femur and the tibia. This will be the focus of future work.

In conclusion, the proposed model based object reconstruction and tracking framework from biplane fluoroscopic sequences represents a great step toward elimination of prior 3D imaging and minimization of manual labor for kinematic studies. The high convergence rate for both estimated and reference standard bone surfaces show the method's potential of minimal user interaction tracking. Also, the tracking precision of about one millimeter and one degree with a statistical shape model is promising for future applications.

3.8   Acknowledgments

Chapter 4

Conditional shape models for cardiac motion estimation

Abstract

We propose a conditional statistical shape model to predict patient specific cardiac motion from the 3D end-diastolic CTA scan. The model is built from 4D CTA sequences by combining atlas based segmentation and 4D registration. Cardiac motion estimation is, for example, relevant in the dynamic alignment of pre-operative CTA data with intra-operative X-ray imaging. Due to a trend towards prospective electrocardiogram gating techniques, 4D imaging data, from which motion information could be extracted, is not commonly available. The prediction of motion from shape information is thus relevant for this purpose. Evaluation of the accuracy of the predicted motion was performed using CTA scans of 50 patients, showing an average accuracy of 1.1 mm.

4.1 Introduction

Coronary angioplasty is an often applied procedure to reopen narrowed or occluded coronary arteries. Real time X-ray visualization guides these interventions, providing information about the morphology of the patent vessels. For difficult interventions, such as reopening chronic total occlusions, integration of a pre-operative CTA acquisition is expected to improve the intervention result. Hereto, a correct alignment of the CTA image with the 2D X-ray projection images is required, which consists of both a rigid alignment to determine the pose and orientation of the data and a non-rigid alignment to compensate for cardiac motion. A patient specific cardiac motion prior can be derived from 4D retrospectively gated CTA data [93]. However, the trend in cardiac CT acquisition is towards prospective electrocardiogram (ECG) gating techniques to decrease the effective patient dose. In these situations, the derivation of dynamic information from CTA data is often not possible, as only one phase of the cardiac cycle is imaged. Moreover, it is often not known before the CTA acquisition if a patient will undergo coronary angioplasty, making selective dynamic imaging for a subset of patients difficult.

The purpose of this work is to develop and evaluate a method to predict patient specific cardiac motion from a single 3D CTA image. To this end a 4D statistical shape model is built from 50 ECG-gated dynamic CTA images by combining 3D multi-atlas registration and 4D registration. Motion is estimated for an unseen patient by first segmenting the cardiac structures in the acquired 3D CTA scan. Subsequently we derive the most probable motion given the segmentation and the statistical model using a conditional Gaussian distribution.

Statistical shape models have been frequently applied for 3D image segmentation [55]. The use of active shape models for dynamic image segmentation has been investigated, for instance in the work of Ordas et al., who built a single 3D model with training data from multiple ECG phases [99]. However, this method cannot be used for the cardiac motion prediction problem, as a single 3D shape model of the heart is built, without distinguishing between inter-patient and intra-patient variability. Models built from 4D landmark positions can overcome this problem. Such a 4D statistical model was, for instance, built by Perperidis et al. for segmenting the left ventricle, right ventricle, and myocardium, but was not used for motion prediction [102]. Hoogendoorn et al. built a bilinear model for the extrapolation of cardiac motion, assuming that the motion of the heart is independent of its shape [58]. We, in contrast, build statistical shape models of the shape and motion of the heart without assuming their independence and evaluate its applicability for the prediction of cardiac motion by conditioning the motion model on single time point shape information.

4.2 Methods

The statistical model is built using multi-atlas based segmentation and 4D registration. The structures of interest are the aorta (Ao), endocardium left ventricle

(endoLV), epicardium left ventricle (epiLV), right ventricle, left atrium (LA) and right atrium (RA). An atlas landmarking procedure ensures anatomical landmark correspondence. Cardiac motion is predicted by conditioning the motion model on the landmark points of the shape in the 3D end-diastolic image. These steps are explained in more detail in the following sections.

4.2.1   Statistical Shape Model

The K 4D CTA sequences that are used to build the statistical models [27] consist of T time points, each of which contains 3D shapes represented by n landmark points. Landmark correspondence is ensured both along the sequence and across the training set (see Sect. 4.2.2). First, a base time point b is determined, which is defined as the time point in the cardiac cycle for which the 3D reconstructions are available. In this work, we use the end-diastolic time point for this purpose. Subsequently, we represent the 4D sequence as a 3D shape in time point b, and a motion sequence given by the vectors describing the displacements from this time point to all time points in the sequence. Accordingly, two separate models are built: a 3D statistical shape model (at time point b) and a 4D statistical motion model. Because the models are created for prediction purposes, the alignment (translation, rotation and isotropic scaling) of the shapes in the training sequences is performed using the landmarks of the shape in the base time point. This is in agreement with the alignment of the segmented end-diastolic target shape to the model. Assuming a Gaussian shape and motion distribution, PCA analysis is performed to extract the main modes of variation resulting in the following shape and motion model

$$\vec{s}_b \approx \bar{\vec{s}}_b + \vec{\Phi}\vec{p}, \qquad \text{and} \quad \vec{m} \approx \begin{bmatrix} \vec{s}_1 - \vec{s}_b \\ \vdots \\ \vec{s}_T - \vec{s}_b \end{bmatrix} = \bar{\vec{m}} + \vec{\Psi}\vec{q}, \qquad (4.1)$$

where $\vec{s}_i$ is the shape of the i-th time point of a sequence, b denotes the base time point, and $\vec{m}$ is the motion vector. The mean shape and motion is denoted with $\bar{\vec{s}}_b$ and $\bar{\vec{m}}$, while $\vec{p}$ and $\vec{q}$ are the shape and motion parameter vectors respectively. The models were reduced to explain 95% of the total variance.

4.2.2   Model Construction

Atlases and Atlas Point Correspondence:

3D segmentation of the cardiac structures at one time point in the cardiac cycle is achieved by multi-atlas registration using $J = 8$ atlases [66]. We used `elastix` for all registration procedures in this work [68]. The following procedure was applied to the atlases to obtain automatic anatomical landmark correspondence:

1. Non-rigid B-spline registration of all atlas pairs to each other, resulting in $J * J$ transformations denoted as $\vec{T}_{ij}$ where i and j indicate the atlas numbers.

2. Determination of the mean transformation for every atlas: $\bar{\vec{T}}_i = \frac{1}{J} \sum_j \vec{T}_{ij}$ and subsequent transformation of the manual annotations to the mean space.

3. Creation of a signed distance map SDM_{is} for every atlas i and structure s in the atlas (a) and mean (m) space and averaging of the signed distance maps for all structures in the mean space: $\overline{\text{SDM}}_s^m = \frac{1}{J} \sum_i \text{SDM}_{is}^m$.

4. Creation of a mean surface with approximately equally distributed landmark locations for every structure by extracting the surface at the zero level set of $\overline{\text{SDM}}_s^m$. The number of landmarks used was 15835 (Ao: 1316, endoLV: 2723, epiLV: 4106, LA: 2034, RA: 2079, RV: 3577).

5. Transformation of the obtained mean surface landmarks back to the atlas spaces using $\bar{\vec{T}}_i^{-1}$ and projection of the landmark points of the mean surfaces onto the annotated surfaces by non-rigidly registering their SDMs.

Landmark Determination in 4D CTA Sequences:

3D segmentation of the heart structures in the end-diastolic time point of the K 4D CTA sequences is performed by:

1. Non-rigid B-spline registration of all atlases to the end-diastolic CTA image.

2. Transformation of the surface landmarks from the J atlases to the target image using the resulting transformations.

3. Combination of the J atlas landmark sets into one landmark set by application of the mean shift algorithm to the landmark coordinates [26]. This ensures a robust average of the corresponding landmark points and reduces the influence of possible registration errors.

This procedure results in a set of landmarks describing the cardiac structures at end-diastole. Subsequently, the segmentation of the cardiac structures in all time phases is achieved by determining the cardiac motion for each patient using 4D B-spline registration on the 4D CTA image. During registration, the variance of intensity values at corresponding spatial locations is minimized. Only deformations in the spatial domain are allowed and the deformation is forced to be periodic.

By combining the segmented shapes and the deformations resulting from the 4D registration procedure we determine 4D landmark sets for every patient. To ensure temporal correspondence between the 4D landmark sets, an additional alignment step is performed to align the moment of contraction and relaxation during the cardiac cycle between patients. To this end, time curves of left ventricular volume are obtained from the K 4D segmentations. These curves are subsequently aligned by a 1D non-rigid group-wise registration approach in which the cross-correlation between the curves is optimized. After this alignment, new landmark positions at twenty regular intervals of the cardiac cycle are determined by applying spline interpolation between corresponding landmark points in time. Volume graphs before and after alignment are shown in Figure 4.1.
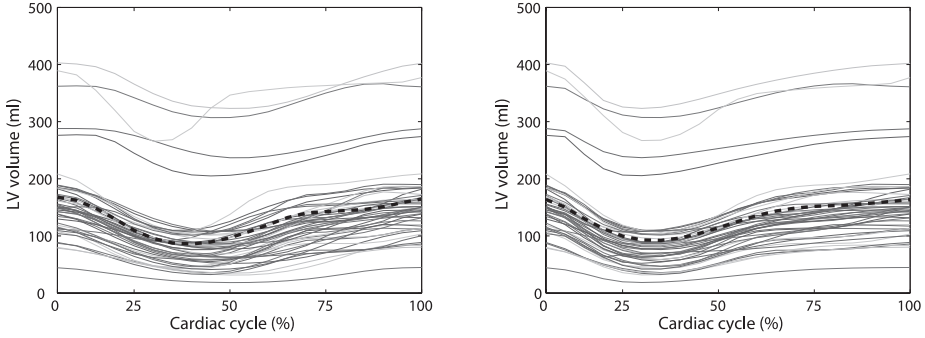
Figure 4.1: Left ventricular volume during the cardiac cycle before and after alignment of the volume curves. Each thin line represents a patient and the dashed black line represents the reference curve.

4.2.3   Conditional Model

We derive the motion given the shape of the cardiac structures at one time point in the cardiac cycle by assuming a Gaussian distribution of the shape vectors $\vec{s}_b \in S$ and motion vectors $\vec{m} \in M$ respectively. Their combined distribution $P(M, S)$ is a normal distribution with mean

$$\vec{\mu} = \left[\begin{array}{c} \vec{\bar{m}} \\ \vec{\bar{s}}_b \end{array} \right] \quad \text{and variance} \quad \vec{\Sigma} = \left[\begin{array}{cc} \vec{\Sigma}_{MM} & \vec{\Sigma}_{MS} \\ \vec{\Sigma}_{SM} & \vec{\Sigma}_{SS} \end{array} \right], \tag{4.2}$$

with $\vec{\Sigma}_{SS}$ and $\vec{\Sigma}_{MM}$ the covariance matrix of shape and motion respectively, and $\vec{\Sigma}_{SM}$ the covariance matrix between shape and motion. The most likely motion $\hat{\vec{m}}$ given a single shape $\vec{s}_b^*$ is estimated as the mean of the conditional probability function $P(M|\vec{s}_b^*)$ [34]:

$$\hat{\vec{m}} = \vec{\bar{m}} + \vec{\Sigma}_{MS} \vec{\Sigma}_{SS}^{-1} (\vec{s}_b^* - \vec{\bar{s}}_b) \tag{4.3}$$

However, as the dimensionality of the 3D shape vectors is larger than the number of training shapes, the covariance matrix $\vec{\Sigma}_{SS}$ becomes singular, and cannot be inverted. Therefore, conditioning is performed on the shape and motion model parameters after applying PCA rather than on the shape and motion vectors themselves. As the mean parameter vectors are zero, Equation 4.3 simplifies to:

$$\hat{\vec{q}} = \vec{\Sigma}_{qp} \vec{\Sigma}_{pp}^{-1} \vec{p}^*, \quad \text{where} \quad \vec{p}^* = \vec{\Phi}^T (\vec{s}_b^* - \vec{\bar{s}}_b), \tag{4.4}$$

with $\hat{\vec{q}}$ being the parameter representation of the estimated motion $\hat{\vec{m}}$. The generation of plausible motion is ensured by scaling $\hat{\vec{q}}$ such that it always lies within ± 3 standard deviations of the motion model. The final conditional sequence is computed from the known 3D shape $\vec{s}_b^*$ and the derived motion $\hat{\vec{m}} = \vec{\bar{m}} + \vec{\Psi} \hat{\vec{q}}$.

4.3 Experiments and Results

4.3.1 Imaging Data

Eight 3D diastolic CTA reconstructions were used as the atlas images and 50 retrospectively gated 4D CTA images were used for building and evaluating the 4D model. All scans were acquired with Siemens CT scanners and 20 time points were reconstructed for the 4D CTA scans. Note that the variety of the data is large as healthy subjects are not imaged because of radiation regulations.

4.3.2 Cardiac Motion Prediction

The accuracy of the motion prediction was evaluated in leave-one-out experiments by applying the conditional model described in Sect. 4.2.3 on end-diastolic segmentations, which are most often available in clinical practice. The accuracy of the resulting 4D shapes was determined by computation of the average root mean squared point-to-surface distance per structure and time point. The results are compared to a projection of the complete sequence onto the model space. This projection represents the best possible reconstruction of an entirely known 4D input sequence given the training data. Note, that this represents a practical limit for accuracy and that the 4D information is not available for the motion estimation. Also, a replication of the end-diastolic segmentations for all time points was used as a worst-case estimate for benchmarking the errors of the conditional shape estimates. Graphs of the accuracy over time are shown in Figure 4.2. Table 4.1 lists the root mean squared point-to-surface distance for both the replication and the conditioning averaged over all time points and over the time point at which the left ventricular volume was minimal (30%). We expect the shapes at this time point to differ the most from the end-diastolic shapes used for the motion estimation. Figure 4.3 shows a color coding of the distances for the left and right ventricle, and atrium at end-systole and end-diastole for two randomly selected patients.

4.3.3 Model Generalization Ability and Training Set Size

An experiment was conducted to assess the dependency between the accuracy of the model and the training set size. The available training sequences were grouped randomly in subgroups, each containing 5 sequences. Subsequently, for every tested training set size all possible combinations of groups were taken as training set, and a random left out group was used for testing. Fitting was performed by projecting the motion of the 4D segmentation of a CTA sequence on the subspace of the motion model, and back to the original space to derive the motion vectors for these parameters. The resulting motion was applied to the atlas segmented shape, and compared with the actual 4D segmentations. Figure 4.4 shows the results.
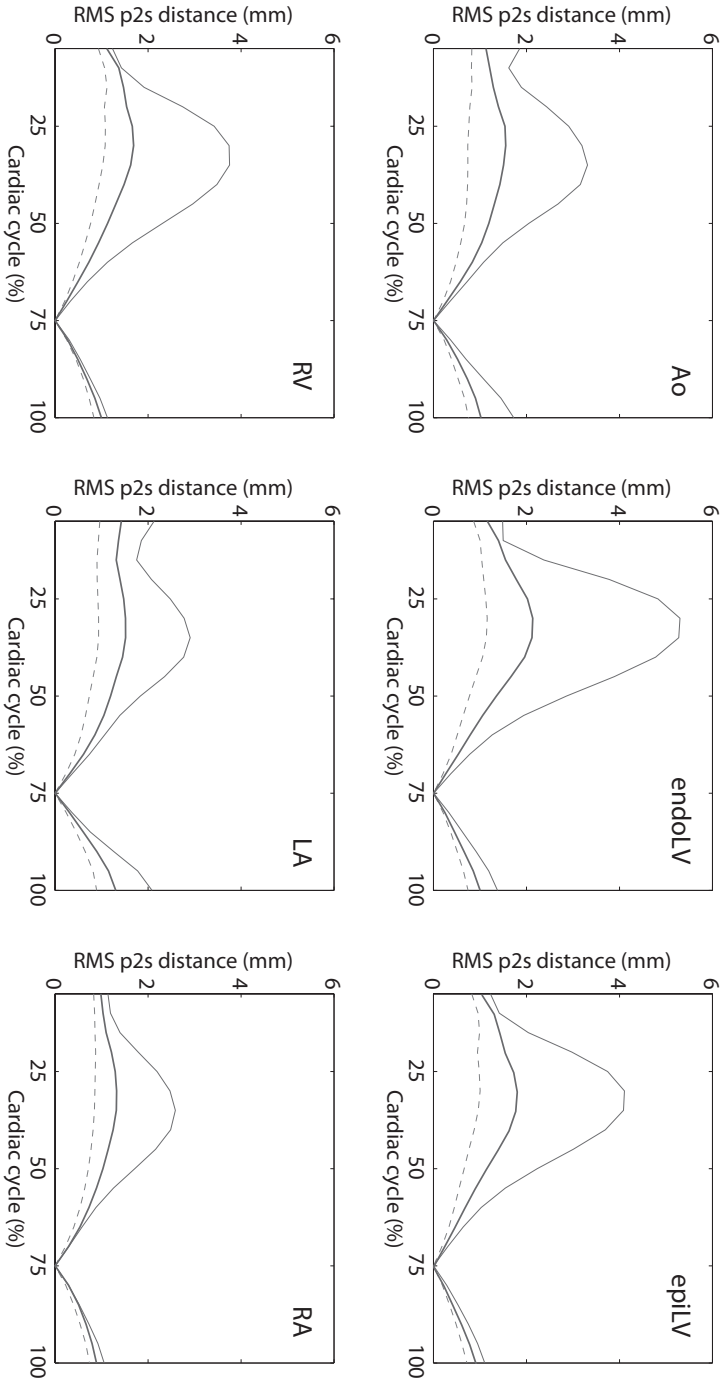
Figure 4.2: Average root mean squared (RMS) point-to-surface (p2s) distance over the cardiac cycle. RMS values are averaged over all patients. The thin line indicates the replication error. The grey dotted line shows the error for projecting the entire sequence onto the model space and the thick continuous line shows the error when conditioning is used.

Table 4.1: Mean and standard deviation of root mean squared point-to-surface distance in mm. for replication (assuming no motion) and the proposed method. Values are listed for all time points and for the minimal left ventricular volume time point (30%).

Structure	Ao	endoLV	epiLV	RV	LA	RA
Replication (all)	1.6 (1.1)	2.2 (1.8)	1.7 (1.4)	1.7 (1.3)	1.6 (1.0)	1.2 (0.9)
Conditioning (all)	0.9 (0.6)	1.1 (0.8)	0.9 (0.6)	0.9 (0.7)	1.0 (0.6)	0.8 (0.5)
Replication (at 30%)	3.2 (1.0)	5.1 (2.0)	4.0 (0.9)	3.7 (0.8)	2.8 (0.8)	2.5 (0.6)
Conditioning (at 30%)	1.4 (0.6)	2.0 (0.7)	1.7 (0.5)	1.5 (0.5)	1.4 (0.5)	1.2 (0.5)
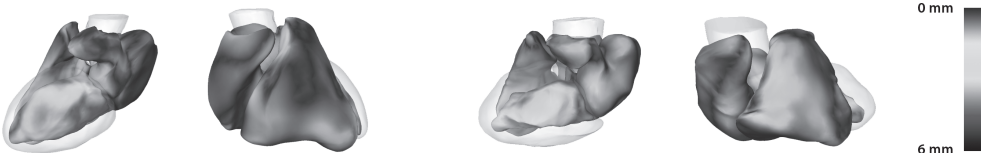
Figure 4.3 (color version on page 156): Two views of end-systolic predicted surfaces for two randomly selected patients. Color coded are the point-to-surface distances from the predicted shape to the segmented shape. The Ao and epiLV are for visualization purposes not color coded.

4.4   Discussion and Conclusion

The experiments show that the conditional shape model is able to predict the motion of the heart with an average accuracy of around 1.0 mm (see Table 4.1). This represents a large improvement compared to shape replication, which is the only option when only a 3D CTA scan is available (see Figure 4.2). Accuracy is measured by the point-to-surface error, but depending on the application of interest, the point-to-point error can be used as well. For the here presented experiments, point-to-point errors were on average larger, but showed very similar trends.

Furthermore, the remaining errors when projecting the entire sequence onto the model subspace and the results of the generalization ability experiment show that the size of the training set is most probably not sufficient. Therefore, the effect of adding more training shapes on the accuracy will be investigated in the future. The errors at the left of the graph are related to the magnitude of the motion found by the registration approach. Errors induced by landmark propagation are currently not taken into account and subject to future work.

In the current work no assumptions were made about the relation between cardiac
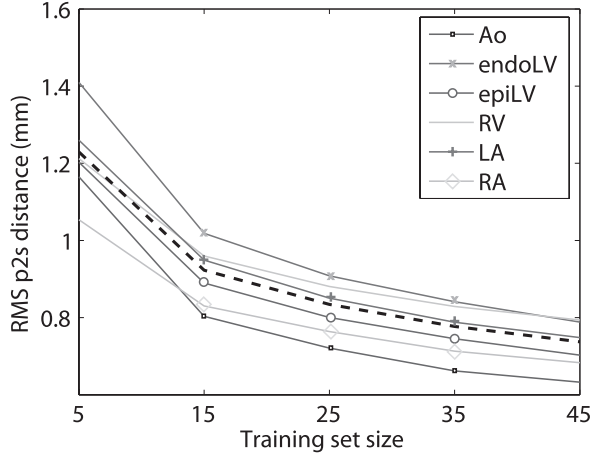
Figure 4.4: Root mean squared point-to-surface (p2s) distance in mm. for every structure with respect to training set size.

shape and motion, in contrast to the method proposed by Hoogendoorn et al., in which shape and motion variations are explicitly decoupled [58]. Although the motion of the heart depends for a large amount on its electrophysiology, which is not apparent from its shape, we think that shape can still predict cardiac motion when both the shape and motion of the heart are affected by disease (e.g. due to cardiac remodelling after myocardial infarction). Motion changes related to the shape at other time points in the cardiac cycle are not taken into account.

The conditioning framework described in this chapter exploits any available correlation between shape and motion given in the training sequences, and naturally reduces to the mean motion in case no correlation is present. Due to the small training set size we were not able to draw any conclusions with respect to this correlation. We plan to investigate this topic in detail as part of future work. For the conditioning 95% of the statistical shape model variance was retained, however the effect of the value of this parameter on the prediction results should still be investigated.

In conclusion, we presented and evaluated a method that uses conditional shape models to predict cardiac motion from 3D cardiac shape. The results suggest that it can reliably predict the motion of the heart. Our goal is to use this model for the dynamic alignment of pre-operatively acquired CTA images with intra-operative X-ray imaging, which should be applied and evaluated in future work.

Chapter 5

Confidence of Model Based Shape Reconstruction from Sparse Data

Abstract

Statistical shape models (SSM) are commonly applied for plausible interpolation of missing data in medical imaging. However, when fitting a shape model to sparse information, many solutions may fit the available data. In this chapter we derive a constrained SSM to fit noisy sparse input landmarks and assign a confidence value to the resulting reconstructed shape. An evaluation study is performed to compare three methods used for sparse SSM fitting w.r.t. specificity, generalization ability, and correctness of estimated confidence limits with an increasing amount of input information. We find that the proposed constrained shape model slightly outperforms the other models, is robust against the selection and amount of sparse information, and indicates the shape confidence well.

5.1 Introduction

Statistical shape models (SSM) are commonly employed in medical image processing for tasks such as segmentation, classification and reconstruction. Their ability to interpolate missing information in a plausible manner, learned from training examples, makes them a favored tool in case of partial data support. Recently, a new application area emerged for sparse SSM fitting: dense 3D shape reconstruction from sparse point clouds, projections, and cross sections [104, 139].

When fitting a shape model to sparse input information, multiple shapes may fulfill the criteria to fit the available data. In a clinical setting, indicating only one of the solutions (even if it is the most probable one) may be misleading, and can lead to wrong conclusions. It is thus important to also indicate the local confidence of the reconstructed shape.

An approach to integrate uncertainties in registration and segmentation has been presented for statistical deformation models [120]. Previous work with point distribution models (PDM) includes empirical evaluation of the effect of data sparsity for dense 3D reconstruction with SSMs [139]. Others indicate the confidence of a shape assuming knowledge of the perfect pose parameters of the model as well as the exact landmark positions of a certain part of the point based shape [3, 16].

These simplifications, however, do not hold in real situations: the recovered pose is dependent on the available sparse information; and the exact knowledge of a landmark point position is not feasible due to partial volume effect and movements, resulting in blurred edges, and the absence of clear anatomical landmarks resulting in a shift in point correspondence.

Directional constraints were introduced for dense shape model fitting to overcome the problem of shift in point correspondence [57]. However, this work does not deal with sparse input, nor does it associate the result with a confidence measure.

In this chapter we present ways to incorporate the data point location uncertainties in the final reconstructed shape confidence for sparse fitting using PDMs. The data point uncertainties may differ from point to point, and are in real situations derived from the underlying image information and/or imaging setup. In this chapter we simulate elongated uncertainties in the direction of the edge tangent, and base shape reconstruction on a varying amount of sparse data for evaluation of the confidence in different sparseness configurations. In Section 5.2 we derive the constraining method for SSMs. Section 5.3 introduces the shape models used in the experiments described in Section 5.4. An evaluation of the method is given in Section 5.5, followed by conclusions in Section 5.6.

5.2 Method

5.2.1 Notations SSM

An SSM is built from N d-dimensional training shapes (segmentations of training images). Subsequently a correspondence algorithm is applied to represent each shape with the same number n of corresponding points, called landmarks. In order to retain

solely shape variations, the shapes are priorly aligned with Procrustes analysis. Each shape can be represented as a point x in the $n * d$ dimensional shape space. The set of training shapes $X = x^1, x^2, ..., x^N$ forms a point cloud in this space. Principal component analysis is applied constructing a model of the form

$$x = \bar{x} + \Phi b + \epsilon \quad \longrightarrow \quad \tilde{x} = \bar{x} + \Phi b \; , \qquad (5.1)$$

where $\bar{x}$ is the mean shape, Φ is the matrix of the main modes of variation, b is the representation of the shape in parameter space, and ϵ is a small error due to the dimension reduction. This model is usually fitted to a new segmentation by minimizing the squared error criterion $E(p, b) = \sum_{i=1}^{n} ||x_i - \tilde{x}_i(p, b)||^2$, where i denotes a landmark point, and $\tilde{x}$ is the fitted shape dependent on pose p and shape parameters b. A common practice of optimization is to first minimize on the pose parameters, followed by the shape parameters.

5.2.2 Sparse pose estimation

When only sparse information is present for fitting, the error criterion changes from summation over all landmarks to summation over the supported landmarks. Pose estimation in this case may yield different results than the full registration, depending on the set of supported points. To account for this pose change, for every sparse fitting a separate sparse aligned model is built. In a sparse aligned model the training shapes are procrustes aligned prior to model building by the same sparse subset of landmarks. The resulting shape model will have larger total variance than the original due to non-optimal pose alignment for the entire shape, but will represent valid shape variations after pose estimation better. In further sections we assume the rigid parameters to be perfectly estimated, and work with the enlarged model for compensation.

5.2.3 Constrained shape model

In this section we derive the constrained shape model. The aim is to fit a PDM to noisy sparse input landmarks as well as possible while retaining a plausible shape. The result is a most probable shape with a normal distribution defining the possible shape variations given the input.

Fitting of SSMs can be formulated in a maximum aposteriori (MAP) framework [30], maximizing probability

$$P(model|data) \propto P(data|model)P(model). \qquad (5.2)$$

In the original SSM model, the term $P(data|model)$ is a Gaussian distribution with covariance ϵI, where ϵ is based on the remaining variation after model instantiation, and I is the identity matrix.

If the exact point positions of the target shape are partially not known, and partially not exactly known, a block-diagonal uncertainty matrix can be built with the block size equaling the point dimensions. Uncertainties of the supported points can

be derived from the underlying image information and/or imaging setup, while non-supported points will have infinite variance indicating no knowledge whatsoever about their position. This additional uncertainty can be incorporated in the $P(data|model)$ term by adding the two uncertainties yielding a normal distribution with covariance $\Sigma_U = \Sigma_{LocUncert} + \epsilon I$. In case a point is exactly known, the uncertainty value goes to the original ϵ. For not constrained, thus unknown, points the uncertainty is infinite, which conveniently gives 0 in the inverse, facilitating the sparse fitting function.

The covariance at the optimum fit can be expressed using the product rule of two normal distributions as

$$\Sigma_C = \beta(\Sigma_U^{-1} + \Sigma_m^{-1})^{-1}. \tag{5.3}$$

The covariance of the model in the full shape space is denoted by Σ_m, and in the reduced parameter space by C_m. Similar notation is used for the constrained covariance Σ_C in the shape space and C_C in parameter space.

Without loss of generality we can assume that the shape vector can be split in two parts $\boldsymbol{x} = (\boldsymbol{x}_a, \boldsymbol{x}_b)^T$, where part a is unknown, while part b is constrained. The covariance matrices can now be written as block matrices

$$\Sigma_U = \begin{bmatrix} \infty & 0 \\ 0 & \Sigma_{Ubb} \end{bmatrix} \;,\; \Sigma_m = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix} \;, \tag{5.4}$$

Using the matrix inversion lemma, and the inversion of block matrices Equation 5.3 can be rewritten as

$$\Sigma_C = \beta \begin{bmatrix} \Sigma_{aa} - \Sigma_{ab}G^{-1}\Sigma_{ba} & \Sigma_{ab} - \Sigma_{ab}G^{-1}\Sigma_{bb} \\ \Sigma_{ba} - \Sigma_{bb}G^{-1}\Sigma_{ba} & \Sigma_{bb} - \Sigma_{bb}G^{-1}\Sigma_{bb} \end{bmatrix}, \tag{5.5}$$

where $G = \Sigma_{bb} + \Sigma_{Ubb} + \lambda I$, with λ being the regularization constant for the case landmark uncertainties are not sufficient for regularization.

The maximization of the aposteriori probability given in Equation 5.2 can be formulated as the minimization of the negative log likelihood function

$$E(\boldsymbol{b}) = (\boldsymbol{x} - \tilde{\boldsymbol{x}}(\boldsymbol{b}))^T \Sigma_U^{-1} (\boldsymbol{x} - \tilde{\boldsymbol{x}}(\boldsymbol{b})) + \boldsymbol{b}^T C_m^{-1} \boldsymbol{b}. \tag{5.6}$$

The optimal shape is derived from the energy in Equation 5.6 by setting the first derivative to 0. The resulting optimal parameter is

$$b_C = C_C \Phi^T \Sigma_U^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}), \tag{5.7}$$

and the corresponding shape

$$\tilde{\boldsymbol{x}}_C = \begin{pmatrix} \bar{\boldsymbol{x}}_a \\ \bar{\boldsymbol{x}}_b \end{pmatrix} + \begin{bmatrix} 0 & \Sigma_{ab}G^{-1} \\ 0 & \Sigma_{bb}G^{-1} \end{bmatrix} \begin{pmatrix} \boldsymbol{x}_a - \bar{\boldsymbol{x}}_a \\ \boldsymbol{x}_b - \bar{\boldsymbol{x}}_b \end{pmatrix}. \tag{5.8}$$

Note that it is not necessary to know $\boldsymbol{x}_a$, because of the corresponding zero values in the premultiplying matrix.

This general formulation produces the conditional model [34] for the covariance as well as the optimal shape in case the constrained points have covariance 0, are thus

exactly known. In this case though, G is likely to be singular, unless the number of training shapes is larger than the number of known point coordinates. A common practice to solve the inversion problem is regularization with a parameter λ [16, 34], which effectively comes down to adding the same isotropic uncertainty to each of the known points. Our approach can also be interpreted as an individual point based regularization, where the regularization parameter is derived from the certainty of a landmark point position. Intuitively, less certain points inherit a larger regularization term. An advantage of this framework is, that the regularization variance can be individualized and directed, for example in the tangential direction of an edge.

5.2.4 Shape confidence

We derive a nominal confidence for each landmark point for the constrained model. We first represent the final variance Σ_C of the fitted SSM in a lower dimensional space, and later compensate for this dimensionality reduction induced error by adding some extra variation to every dimension in the shape space prior to confidence calculation. The α confidence interval is then calculated for each individual landmark, meaning that the landmark is with a probability of α inside the indicated confidence area. The delineating Mahalanobis distance D_α is derived from $D_\alpha^2 = K^{-1}(\alpha, d)$, where K^{-1} is the inverse chi square distribution with dimension d (we refer to [16] for a derivation).

5.3 Data

Experiments were carried out on two PDMs:

(i) A 2D left ventricle (LV) model, made from 50 mid cross sectional MRI slices of healthy volunteers. Epi- and Endocardium are both represented by 40 points sampled equidistantly on the manually segmented contour, starting by one end of the interventricular septum.

(ii) A 2D left and right ventricle (LVRV) model, built from 114 MRI slices out of the available 125 after excluding 11 outlier shapes. The data was collected from 42 healthy and diseased hearts. Image acquisition and population is described in detail in [95]. Point correspondence is defined by equidistant sampling of the right ventricle, the interventricular septum, and the rest of the Endo- and Epicardial contours separately, resulting in total 105 landmark points. When referred to leave one out (LOO) tests in context of this model, all slices of the same patient are left out of the training set.

5.4 Experiments

We compare three models: 1) the constrained shape model as introduced in Section 5.2; 2) the conditional shape model with optimal regularization parameter (ridge regression); and 3) the same conditional shape model with added variation. The conditional model treats the target points as exact locations, and gives the most probable fit. One can argue, that a slightly different input will produce a slightly

different predicted shape. Generally, a $\Sigma_{LocUncert}$ variance in the input points will produce a

$\Sigma_{extra} = \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{LocUncert}\Sigma_{bb}^{-1}\Sigma_{ab}^T$ variance in the fitted landmark positions. The conditional model enlarged by this extra variation has the same mean shape, but larger covariance than the standard conditional model. For completeness reasons, we compared the constrained method with both conditional models.

Experiments were carried out in a leave one out fashion on the LV and LVRV models due to the small number of training shapes. In order to assess the effect of sparsity on the reconstruction, an increasing number of points were selected as 'known'. As close points are likely to be correlated, we evaluated the fit for neighbouring and random sparse selections. Point uncertainties are simulated by adding noise to the known point locations, mainly in the tangential direction. The same generating variance was used as the uncertainty matrix $\Sigma_{LocUncert}$ for constraining as well as for deriving the optimal regularization parameter for models 1) and 2). The parameter was calculated by leave one out cross validation for each sparsity scenario with random noise drawn from the simulated uncertainty distribution. The regularization parameter estimation of model 3) did not involve uncertainty estimates.

For comparison the following measures were computed.

- *Specificity*, indicating how plausible the remaining variations are after fitting. 200 random shapes were generated from the fitted model within the $\pm 3\sigma$ hyperbox. The specificity is the percentage of generated shapes that lie within the $\pm 3\sigma$ hyperbox of the original model, created with prior training shape alignment using all landmark points.

- *Generalization*, indicating how well the correct shape could be predicted by the sparse fit. It is calculated as the root mean square error between the landmarks of the real shape without the simulated variance and the most probable shape of the fitted model.

- The *effective confidence* measure of Blanc et al. was adopted to validate the confidence regions [16]. The nominal confidence is calculated according Section 5.2.4. The effective confidence value for these regions is calculated as the percentage of landmarks of the real shape that lie within these respective confidence regions. It is desired that the effective and the nominal confidence coincides, or the nominal slightly overestimates the effective value.

- The *total variance* of the fitted model, serving as an indication of its size. It is calculated as the sum of the eigen variances.

5.5 Results and Evaluation

Results on the LV and LVRV models show that the confidence regions calculated from the conditional and the constrained model slightly underestimate the variability of landmark points (Figure 5.1, lower left plot). The larger the number of supported

points, the larger the difference between effective and nominal confidence. The additional variance in model 3) can compensate for this behavior, resulting however in a less specific, and larger model than either constrained or conditional. The constrained model generalizes better than the standard conditional model. Figure 5.1 shows the generalization, specificity, and effective confidence measures for an increasing number of input points on the LVRV model. The sparse input points were selected randomly, while ensuring that all three main parts: right ventricle, endocardium, and epicardium had selected points. Similar results can be obtained for neighboring supported points,
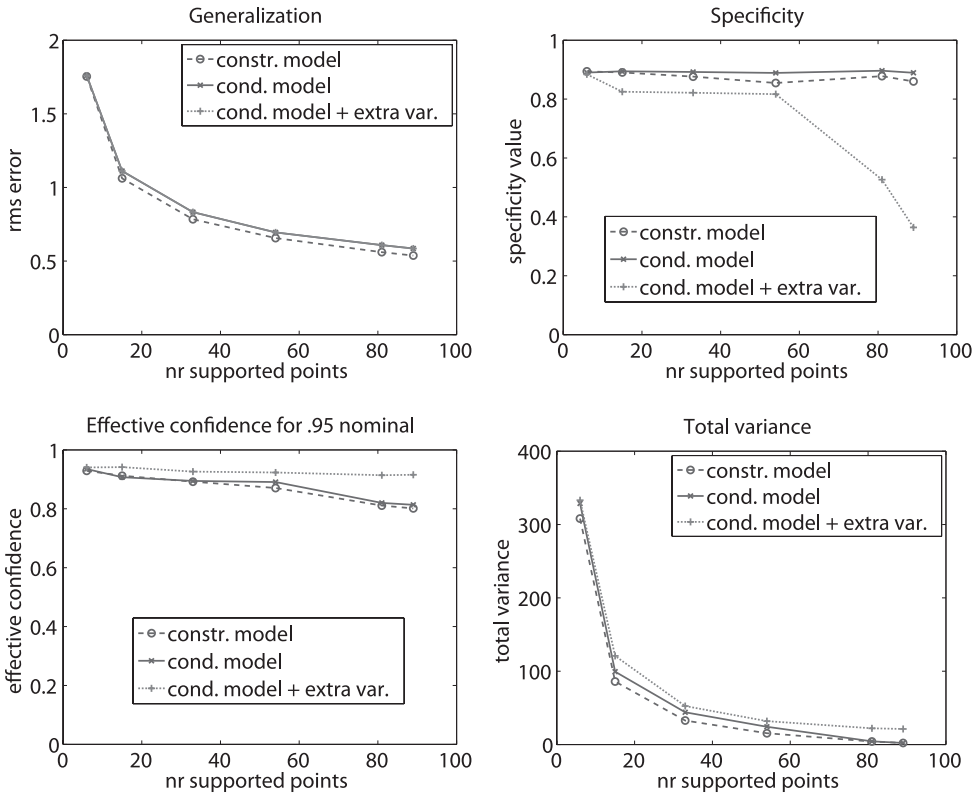
Figure 5.1: Generalization, specificity, effective confidence, and total variance in the model for increasing number of randomly selected input points on the LVRV model

with the difference, that overall errors are larger owing to poor alignment prior to model building, and the relative differences between the models are smaller. Figure 5.2 illustrates the confidence intervals for a particular sparseness scenario, with half of the right ventricle and the endocardial contour being supported.
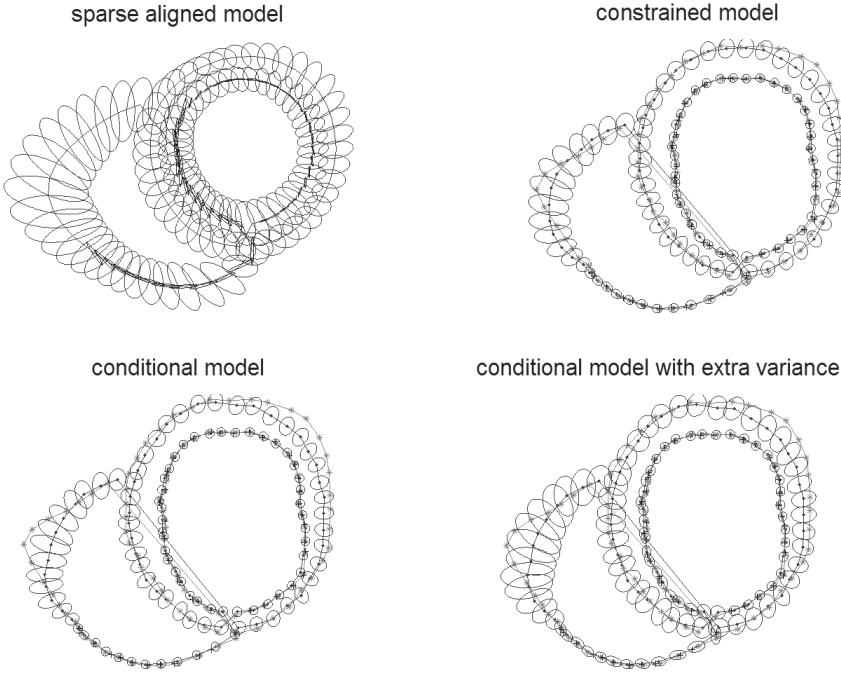
Figure 5.2 (color version on page 156): Confidence intervals of $\alpha = 0.95$ are indicated as blue ellipses (from left to right) on the sparse aligned model, the constrained model, the conditional model, and the conditional model with extra variance. The true shape is drawn magenta with landmarks indicated as *-s. Blue is the reconstruction, based on the black points +, which are noisy representations of the selected landmarks.

5.6 Discussion and Future Work

This chapter dealt with statistical shape instantiation based on sparse data support. The proposed framework allows uncertainties in the target landmark positions. We compared this model with two variants of conditional models (using ridge regression) with respect to generalization, specificity, and effective confidence on an increasing number of supported points with simulated noise. We found that the constrained model predicted the shape the best, and described the remaining variances well, i.e. it resulted in a similar effective confidence and specificity as the conditional model, with a lower amount of total variance. Computed confidences show a slight underestimation of the confidence regions on two 2D heart datasets, for all confidence levels. All three models were built on sparse aligned training shapes. Future work might include pose uncertainties in the final confidence measure. Also, we assume known point correspondences prior to fitting. An interesting generalization would be

confidence of a fit to non-corresponding sparse points. Resulting confidences can be used as evaluation of the sparse information the SSM is fitted on, and selection of the optimal set of sparse information.

Chapter 6

Comparison of shape regression methods under landmark position uncertainty

Abstract

Despite the growing interest in regression based shape estimation, no study has yet systematically compared different regression methods for shape estimation. We aimed to fill this gap by comparing linear regression methods with a special focus on shapes with landmark position uncertainties. We investigate two scenarios: In the first, the uncertainty of the landmark positions was similar in the training and test dataset, whereas in the second the uncertainty of the training and test data were different. Both scenarios were tested on simulated data and on statistical models of the left ventricle estimating the end-systolic shape from end-diastole with landmark uncertainties derived from the segmentation process, and of the femur estimating the 3D shape from one projection with landmark uncertainties derived from the imaging setup. Results show that in the first scenario linear regression methods tend to perform similar. In the second scenario including estimates of the test shape landmark uncertainty in the regression improved results.

6.1   Introduction

Shape estimation by linear regression has been applied in a great variety of problems, including 3D shape estimation from digitized point cloud [104] and projection images [139], neighboring shape prediction [77, 105, 133], organ motion prediction based on a few time-points [2, 69, 92], healthy shape prediction for disease quantification [34], and remaining shape variation prediction from partial field of view [5, 16].

The shapes used to construct the regression model are segmentations of images with noise, artifacts, low contrast, etc. Therefore, landmark positions always inhabit a certain amount uncertainty, which may be estimated from the images, imaging setup, segmentation method, etc. Regression methods that are currently applied for shape prediction do not incorporate this additional information.

Studies in chemometrics literature compared the performance of linear regression methods in noisy environments concluding that including data uncertainty may improve prediction in certain situations [107]. These conclusions can not directly be generalized to shape regression, as they apply to single variable outcome cases with low dimensional input and a large training set. Shape regression on the other hand is characterized by high dimensional input and output variables, and a very small training set.

Therefore, we addressed the following questions: (1) Is there a preferred linear regression method for shape regression? (2) In what context is it beneficial to incorporate landmark position uncertainties in the regression?

We answer these questions by comparing several linear regression methods with and without incorporation of uncertainties. We differentiate two scenarios: In the first, uncertainties of the landmark positions are similar in the training and test dataset, whereas in the second uncertainties of the training and test data are different. We evaluate the performances in simulation experiments providing a controlled setup with a Gaussian shape and uncertainty model and known ground truth, and in a real-world dataset for each scenario. First, the end-systolic left ventricle shape is predicted from end-diastole with landmark uncertainties derived from the segmentation process, and second, the 3D femur shape is estimated from one projection with landmark uncertainties derived from the imaging setup.

6.2   Shape regression

Shape representation.

Each training shape consists of landmark points, which are in correspondence across all shapes. After rigid alignment each shape can be represented by its concatenated landmark coordinates as a high dimensional point. The set of training shapes in this high dimensional space is commonly assumed to be Gaussian distributed with lower intrinsic dimensionality.

Regression techniques.

In this chapter we describe shape regression where both input x and output y are shapes. We focus on linear regression methods, as in low sample size high dimensionality problems non-linear methods tend to over-fit the training data [54]. Ordinary least squares (OLS) is the best unbiased linear estimator. It minimizes the sum of squares prediction error requiring an over-determined system, and often inhabiting large prediction variance[1]. Due to the small sample size and the lower intrinsic dimensionality shape regression is typically under-determined, and OLS is not directly applicable. We compared biased methods as they both reduce prediction variance, and are applicable for under-determined problems.

Standard biased regressions not including landmark uncertainty are principal component regression (PCR) applying OLS on a sub-space calculated by principal component analysis; partial least squares regression (PLS) optimizing the underlying lower dimensional spaces of input and output to maintain the most covariance between them[2]; and ridge regression (RR) applying a norm constraint on the regression coefficients. See [54] for more detailed descriptions.

Maximum likelihood (ML) PCR includes uncertainties of both test and training data by performing OLS on the input subspace A, that is optimal in the ML sense for a given set of samples with known Gaussian uncertainties. A sample x with uncertainty Ψ is represented as $r = (A^T \Psi^{-1} A)^{-1} A^T \Psi^{-1} x$ in this subspace. We defined a minimum uncertainty for all dimensions to maintain stability. Furthermore to cut computational costs, we used an efficient alternating algorithm with independent noise approximation and no intercept optimization [131]. MLPCR has not been used in shape regression before.

To include uncertainty in the test sample only, we used maximum a-posteriori (MAP) ridge regression, proposed in [5]. With known test uncertainty Ψ one obtains the coefficient matrix $B_{\mathrm{MAPR}} = \Sigma_{YX}(\Sigma_{XX} + \Psi + \lambda I)^{-1}$, where Σ_{XX} and Σ_{YX} are the input covariance and cross-covariance matrices, I is the identity matrix, and λ is a regularization constant.

Suitable values for the percentage of variance retained with PCR, number of modes of PLS and MLPCR, and regularization constant for ridge and MAP ridge regressions have to be assessed from the training set. We used cross-validation on the training set to estimate the regularization.

6.3   Simulation study

Framework description.

To evaluate shape regression with different noise patterns, we simulated data that resembles real world shapes, i.e. (1) the data forms a low-dimensional Gaussian distribution with variance outside this subspace being Gaussian noise; (2) noise of neighboring dimensions is correlated; (3) the variance of modes drops quickly and

---

[1] With multivariate x and y OLS is termed multiple outcome multiple linear regression
[2] We used the SIMPLS algorithm for PLS

then smoothly converges to a constant; (4) sample dimensionality greatly exceeds the training set size; and (5) the input only partially predicts the output.

The data was generated with intercepts $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ as follows

$$\boldsymbol{x}^i = \boldsymbol{x}_0 + \Phi_{in}\boldsymbol{\alpha}^i + \boldsymbol{\delta}_{in}^i \quad , \quad \boldsymbol{y}^i = \boldsymbol{y}_0 + \Phi_{out}(W\boldsymbol{\alpha}^i) + \boldsymbol{\delta}_{out}^i \tag{6.1}$$

where $\boldsymbol{\alpha}^i$ is the d_{in} dimensional representation of the i-th training input $\boldsymbol{x}^i$, and is sampled per dimension from a Gaussian distribution with decreasing variance. The matrices Φ_{in} and Φ_{out} contain random orthonormal vectors, and W is the random orthogonal linear regression coefficient matrix of size $d_{in} \times d_{out}$. The noise $\boldsymbol{\delta}_{in}^i$ and $\boldsymbol{\delta}_{out}^i$ is sampled from a Gaussian distribution with covariance

$$\Sigma_\delta = \zeta\left[\Sigma_{common} + \Sigma_{sampleSpec}\right] = \zeta\left[\Theta^T\Gamma\Theta + S^TS\right] \quad , \tag{6.2}$$

where ζ regulates the amount of noise. Noise consists of Σ_{common} producing a shared noise structure, and a sample specific term $\Sigma_{sampleSpec}$. The first is calculated from a random low dimensional space spanned by the columns of Θ with a magnitude defined by the diagonal of Γ, while the second is calculated from a random matrix S consisting of uniform gaussian samples, except of a diagonal band exhibiting larger variance. Such noise models the common uncertainty structure shared by all samples, such as a larger error along the shape surface than perpendicular to it as well as the smoothness of real shapes making noise highly correlated in a neighborhood.

Experiments and results

We conducted experiments with 200 dimensional input and output of intrinsic dimensionality $d_{in} = 5$, $d_{out} = 3$. We chose ζ such that the noise magnitude was 10% of the real data variation. Training set size was 30, unless stated differently. For the first two experiments the exact noise covariance was used as sample uncertainty estimate. We used the root mean squared (RMS) distance of the predicted output and the noise-free test output for evaluation. The reported results are averages of 100 random realizations.

The first experiment simulates applications where training and test shapes are produced with the same segmentation method and imaging protocol. This scenario is most common in real shape regressions. For this simulation training and test input and output is given same amount of noise. Results for different training set sizes shown in Figure 6.1 indicate only small differences between regression methods. Results of 2-sample t-tests on the pooled training set sizes are shown in Table 6.1. The same experiment with 50% noise gave similar results.

The second experiment simulates applications where the training shapes are created from a different modality or with a different method than the test data, e.g. training shapes from CT and test data from X-ray [139]. They exhibit thus different noise structure. We simulated the training set with 10% noise, while test sample noise varied between 5% and 100%. Results are shown in Figure 6.2 and Table 6.1.

In the third experiment we assessed the effect of wrongly estimated amount of test uncertainty. We used the setup of the previous experiment with 10% training set
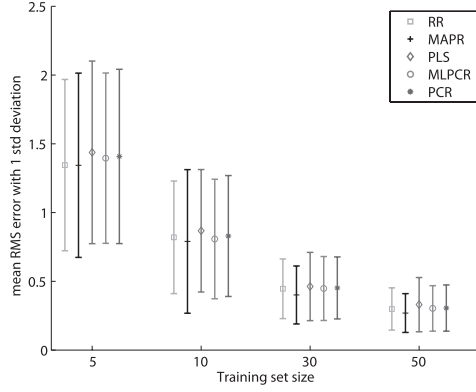
Figure 6.1: Error as function of training set size for different regression methods, when noise in training and test data is similar (Simulation experiment 1)

		S1	H	S1	H	S1	H	S1	H
RR		↓	0	←	←	0	←	←	←
→	→	MAPR		←	←	←	←	←	←
↑	↑	↑	↑	PLS		↓	0	↓	↓
0	↑	↑	↑	→	↑	MLPCR		0	↓
0	↑	↑	↑	→	0	0	→	PCR	
S2	F	S2	F	S2	F	S2	F		

Table 6.1: Significance table: The upper triangle summarizes results when training and test data are similar (S1 - simulation 1, H - Heart). The lower triangle summarizes experiments with different training and test data (S2 - simulation 2, F - femur 2D-3D). The arrows point to the significantly outperforming method (5% confidence level).

noise, and twice as large test noise, and varied the estimated magnitude of the test uncertainty. Note, that the covariance structure of the test uncertainty stayed the same, only the magnitude varied. Results are shown in Figure 6.3.

6.4   Real world datasets

Left ventricle prediction

Estimation of a patient-specific 4D beating heart model is useful for integration of pre-operative 3D CTA in X-ray guided coronary angioplasty procedures [92]. We investigated the accuracy of predicting the end-systolic left ventricle shape from end-diastole as a real world example for the scenario of similar test and train uncertainties.

We extracted 150 end-diastolic and end-systolic left ventricle (LV) shapes from 4D cardiac CTA segmentations, derived by atlas segmentation at end-diastole and subsequent 4D registration as proposed by [92]. The atlases contained 5899 landmark
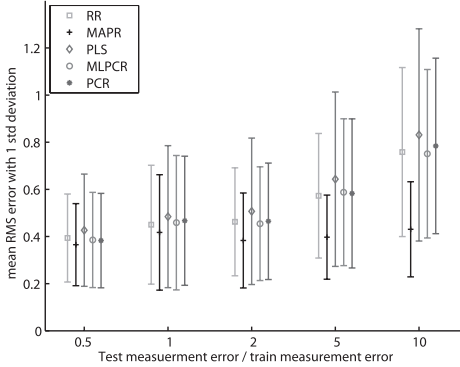
Figure 6.2: RMS error for different regression methods as function of the ratio between test and training set noise magnitude (Simulation experiment 2)
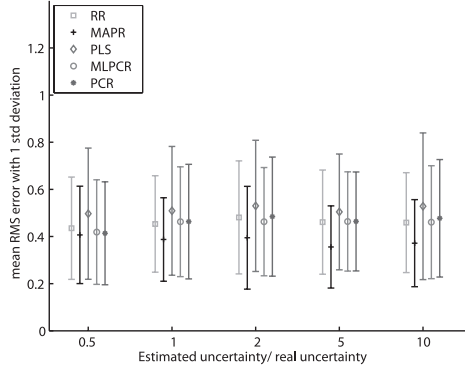
Figure 6.3: RMS error for different regression methods as function of the ratio between estimated and actual test noise magnitude (Sim. experiment 3)

points denoting the LV endo- and epi-cardium. The end-diastolic segmentations were obtained by averaging the landmark points of the eight registered atlases. The spread of these points was used to estimate the landmark uncertainty. Example segmentations with color-coded uncertainty magnitude per landmark are shown in Figure 6.4.

The RMS point-to-point (P2P) distance between predicted and segmented end-systolic shape was evaluated in leave-one-out experiments. The regularization parameters in PCR, PLS and RR were optimized via 15-fold cross-validation, and MLPCR was assigned the same number of modes as PCR. Due to the high dimensional input, MAPR was performed in the data subspace, and was optimized per test sample. Results for the different regression methods are shown in Figure 6.6 with Table 6.1 indicating statistical significance.

2D-3D femur reconstruction

Estimation of 3D patient specific bony anatomy from one or more X-Ray images would be beneficial to minimize acquisition costs and radiation dose. Regression based reconstruction after rigid alignment and 2D-3D correspondence generation was proposed by Zheng et al. [139]. We focused on the reconstruction step assuming known projection parameters, optimal pose, and perfect 2D-3D correspondence. This is an example of the scenario when training and test uncertainty differs.

The regression coefficient matrix was calculated with as input the 3D silhouette landmarks and as output the entire 3D shape. The test input was generated from the 2D projected silhouette landmarks as the 3D point on the projection ray, which is closest to its corresponding landmark of the mean shape. Therefore, the test input was given a large uncertainty along the projection rays, and a small one perpendicular to it, as shown in Figure 6.5. The training shapes were rigidly aligned based on the
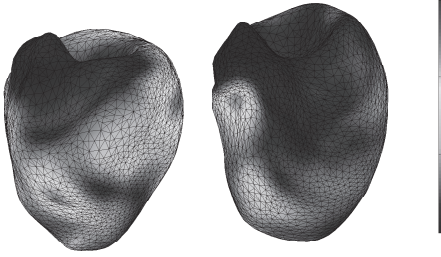
Figure 6.4 (color version on page 157): Two end-diastolic left ventricle shapes with color-coded landmark uncertainty [0-15 mm], derived from CTA data by multi-atlas segmentation.
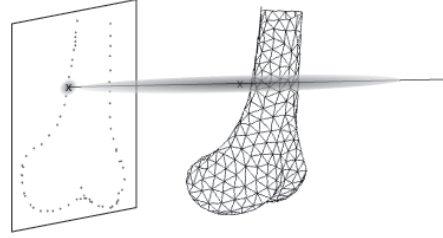
Figure 6.5: The 2D-3D uncertainty structure. The 3D position of a projected 2D silhouette point is highly uncertain in the projection direction.
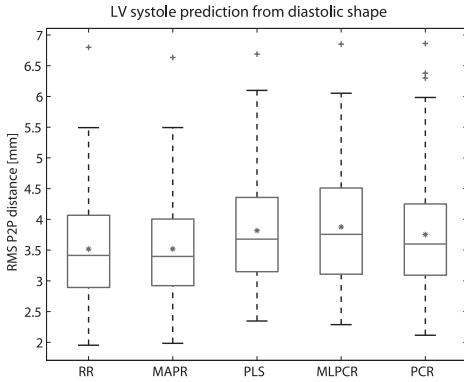
Figure 6.6: Box plots showing RMS errors between estimated and reference LV systolic shape. The blue star marks the mean RMS error.
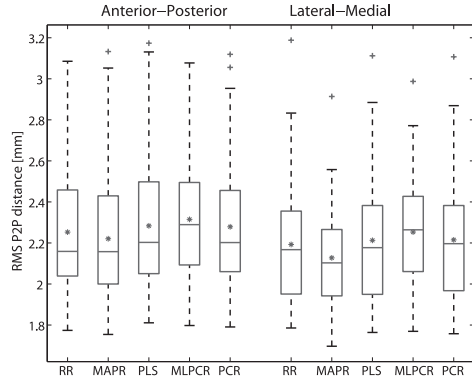
Figure 6.7: Box plots showing RMS errors between estimated and reference 3D femur shape for two projection directions.

silhouette landmark points only, as proposed in [5]. We performed leave-one-out experiments to estimate the 3D distal femur shape from one projection.

The femur model was created from semi-automatic segmentations of 29 patient CTs and 13 cadavers from both sexes with varying ages. Point correspondence was generated with the GAMES algorithm [42]. The resulting shape model contained 33 modes of variations with 95% retained variance.

We compared the regression methods based on the RMS P2P distance between predicted and test output. Results for the Anterior-Posterior (AP) and the Lateral-Medial (LM) projection directions are shown in Figure 6.7 with significance indication

in Table 6.1.

6.5 Discussion and Conclusions

Despite the growing interest in regression based shape prediction, no study had before systematically compared different regression methods for shape analysis. We aimed to fill this gap by comparing several linear regression methods with a special focus on landmark position uncertainties. We defined two uncertainty scenarios.

In the first scenario test and train uncertainties were similar. Simulation experiment 1 and the LV prediction experiment (Figure 6.1, Figure 6.6, and Table 6.1) showed little differences between different linear regression methods. Uncertainty incorporation via MLPCR did not improve results. A similar phenomenon was reported in [107] for MLPCR without intercept optimization. Results might improve without the simplifications of independent noise and no intercept, but the high computational requirements make it currently impractical for shape regression. Including uncertainty of the test sample via MAPR marginally improved predictions in the simulations, but did not improve upon RR on the heart data. While the regularization constant of MAPR depends on the test uncertainty, and has to be re-optimized for every new test sample, the optimization of the standard regression methods is solely based on the training data, can thus be precomputed. We observed that the average cross-validation curve used to chose the shrinkage parameter was smooth for RR, and spiky for PCR and PLS due to the discrete inclusion of directions, and the false correlation magnitudes in case of a small and noisy training set. This might be one reason why RR slightly but consistently outperformed PCR and PLS. Overall, in this scenario we found RR to perform slightly better than to other linear regression methods.

In the second scenario training and test uncertainties differ largely. Simulation experiment 2 and the 2D-3D femur reconstruction experiment (Figure 6.2, Figure 6.7 and Table 6.1) consistently showed that including knowledge of the test uncertainty in the prediction via MAPR significantly improves results. The practical gain depends on the difference in uncertainty and on how well the assumed normal distribution fits the shape variability and the noise. Real world data might therefore produce a smaller gain with MAPR. Estimation errors of the noise magnitude with the correct covariance structure are, however, tolerable within a large interval (simulation 3, Figure 6.3). The performance of RR, PLS, PCR and MLPCR did not change from scenario 1. Therefore, we conclude that in such situations including landmark uncertainty of the test sample may be beneficial.

In our comparison the predicted shape is the end result. Our work could be naturally extended to include the remaining variance of the prediction.

We believe that the presented comparison study gives insights for researchers working on shape prediction in various domains, and helps in the selection of a suitable regression method for future applications.

Chapter 7

Statistical coronary motion models for 2D+t/3D registration of X-ray coronary angiography and CTA

Abstract

Accurate alignment of intra-operative X-ray coronary angiography (XA) and pre-operative cardiac CT angiography (CTA) may improve procedural success rates of minimally invasive coronary interventions for patients with chronic total occlusions. It was previously shown that incorporating patient specific coronary motion extracted from 4D CTA increases this robustness of the alignment. However, pre-operative CTA is often acquired with gating at end-diastole, in which case patient specific motion is not available.

For such cases, we investigate the possibility of using population based coronary motion models to provide constraints for the 2D+t/3D registration. We propose a methodology for building statistical motion models of the coronary arteries from a training population of 4D CTA datasets. We compare the 2D+t/3D registration performance of the proposed statistical models with other motion estimates, including the patient specific motion extracted from 4D CTA, the mean motion of a population, and the predicted motion based on the cardiac shape.

The coronary motion models, constructed on a training set of 150 patients, had a generalization accuracy of 1 mm root mean square point-to-point distance. Their 2D+t/3D registration accuracy on one cardiac cycle of 12 monoplane XA sequences was similar to, if not better than, the 4D CTA based motion, irrespective to which respiratory model and which feature based 2D/3D distance metric we used. The resulting model based coronary motion estimate showed good applicability for registration of a subsequent cardiac cycle.

7.1 Introduction

Percutaneous coronary intervention (PCI) is routinely performed for treating stenosed coronary arteries. During the procedure the site of lumenal narrowing is dilated and a stent is placed under X-ray coronary angiographic (XA) guidance. The XA sequence shows both the guide wire and the vessels due to the injection of contrast agent. PCI on non-occluded lesions has high success rates (99%). These rates drop to 60-80% in the case of treatment of chronic total occlusions (CTOs) [37, 115]. CTOs can not be fully visualized with XA as the contrast agent does not pass the occluded segment, which hampers percutaneous treatment. In order to successfully cross a CTO with a guide wire, the interventional cardiologist ideally requires visualization of the occluded segment including the vessel borders and plaque composition, because soft parts of the lesion are easier to cross than calcified parts. Pre-operative CT angiography (CTA) can be used to visualize the entire cardiac vasculature including occluded segments, and the possibility to distinguish calcified plaque [83]. The combined visualization of the information from CTA and XA during percutaneous treatment of CTOs may increase the success rate of CTO treatment. Therefore, this chapter deals with the registration of intra-operative XA sequences with pre-operative 3D CTA images. To account for cardiac motion throughout the registration, we investigate the possibility of employing statistical motion models of the coronary artery.

Many authors have addressed registration of a 3D vascular tree with its projection images. While calibrated biplane acquisitions enable direct reconstruction of the centerline structure [97], often only a monoplane setup is available. Registration of the 3D vascular tree with one projection image has been investigated by several authors. Many of the works focus on anatomies which are inherently not moving, such as the cerebral arteries [40, 67], or the coronary arteries using an XA frame taken at the same cardiac phase as the CTA image [91, 108, 124]. However, selecting the XA frame with the correct cardiac phase is not easy, and an inexact selection may deteriorate registration performance [124].

To account for slight differences in the vessel shapes, non-rigid 2D-3D alignment techniques have been proposed. Groher et al. allowed non-rigid deformation of vessel segments of the liver while constraining the length and smoothness of the vessel to ensure plausible deformations [51]. Gatta et al. corrected for shape differences using 2D non-rigid registration between the projected coronary arteries and one XA frame [50]. Serradell et al. generated a synthetic motion model of the coronary arteries to constrain the non-rigid motion while fitting the 3D vessel tree to the 2D projection image [111].

Registration of an entire cardiac cycle rather than a single time-frame provides more information to guide the alignment, but it requires coping with the cardiac and respiratory motion of the coronary arteries. Bouattour et al. performed frame-by-frame B-spline based 2D-3D fitting, with the previous frame's fit as the initialization for the next frame[20]. They reported a success rate of 56% on the evaluated three sequences. Metz et al. proposed simultaneous registration of all frames of a cardiac cycle, using the cardiac motion derived from 4D CTA data [90]. They showed an increase in registration robustness compared to the single frame 2D-3D fitting of

the end-diastolic XA frame. However, in clinical practice CTA images are typically acquired within a limited part of the cardiac cycle in order to minimize radiation dose. Therefore, information on the motion of the coronaries is often not available. Statistical motion models might in this case replace the patient specific motion.

Statistical motion models have been proposed for respiratory motion tracking in image guided radiotherapy of the liver [13] and the lungs [78], in image guided interventions [69], and for modeling heart deformation due to respiration [87]. These models were built from multiple images of a single patient to represent the intra-patient variability of breathing. If only one time-point in the cardiac cycle of the patient is known, construction and application of such models is not feasible.

Statistical motion models have also been proposed to represent the motion variation in a population, thus modeling the inter-patient variability. The straightforward extension of statistical shape models to incorporate motion is by concatenating the landmark coordinates of every frame to derive the modeled coordinate vector. Such 4D shape models have been proposed for dynamic segmentation, such as for left ventricle segmentation from ultrasound images [19], and segmentation of the entire heart from 4D CTA [99] and MRI images [136]. Others solely modeled motion, thus the difference in point coordinates from a reference frame. Such models have been proposed for recognizing pathologies [23, 102, 117], and for predicting cardiac contraction patterns either for the entire cardiac cycle, or a few phases ahead [58, 88].

Coronary artery geometry has previously been studied by Lorenz et al., who built a mean end-diastolic heart surface and coronary artery tree model [80]. Due to the topological variation between patients, the coronary artery model was restricted to the largest arteries, which have a relatively well defined position. A synthetic coronary artery motion model was proposed by Serradell et al., based on the 3D vessel centerlines [111]. The motion model was generated by hierarchical simulated random movements of the centerline points, where each movement of a point affected the entire sub-tree below that point. The resulting movements remained realistic by smoothing and by constraining movements in the vessel direction. To our knowledge, no population based statistical coronary motion models have been reported in literature. This might be due to the high variability in coronary artery anatomy, and the difficulty to reliably segment coronary arteries in a large number of 4D datasets.

In this chapter, we perform 2D+t/3D registration of single-plane XA sequences with pre-operative end-diastolic CTA images. We turn this registration into a 2D+t/3D+t registration by employing a coronary artery motion prior. The primary focus of the chapter is the question whether the coronary artery motion prior needs to be patient specific, i.e. derived from a 4D CTA scan (as proposed in [90]), or could other sources be used, such as a statistical motion model or the mean motion over a population. To investigate this, we propose a methodology to construct patient specific statistical coronary motion models based on 4D CTA images of a training population. The aforementioned problems with population based coronary artery motion models are circumvented by estimating the coronary motion from the motion of the cardiac surface. Four different population based coronary artery motion models are proposed and evaluated for 2D+t/3D alignment: the mean motion, a statistical motion model, the most probable motion based on cardiac shape, and a shape conditional motion

model.

The main contributions of this chapter are the following:

1. We propose a technique for constructing statistical coronary artery motion models from 4D CTA datasets by coronary motion estimation from the cardiac surface, and generate two types of models: a classic statistical motion model and a shape conditional motion model.

2. We investigate if inclusion of the derived motion models can lead to a similar 2D+t/3D registration accuracy as the 2D+t/3D+t registration employing a 4D CTA based patient specific motion model.

Furthermore, we investigate if the motion estimated with a statistical model can be applied in registration of subsequent cardiac cycles. This is of interest, as a patient specific motion prior recovered from high contrast frames could then be used in subsequent cycles with fading contrast.

The chapter is structured as follows. First, we describe the construction of the proposed statistical coronary motion models (Section 7.2). Then, in Section 7.3, we introduce the feature based 2D-3D distance metric used in the registration, and describe the strategy of incorporating different models for the whole cardiac cycle registration. After discussing implementation details in Section 7.4, and the data in Section 7.5, the experiments and results are presented in Section 7.6. Finally, we discuss the results, and present the conclusions in Section 7.7.

7.2 Statistical coronary motion model construction

Building a statistical motion model directly from a training set of 4D coronary artery centerlines is very challenging due to the variations in coronary topology. Therefore, we propose to build a model based on the motion of nearby cardiac regions. The method consists of four steps: 1) Building a training set of 4D heart surfaces with point correspondence, 2) Segmentation of the cardiac structures and the coronary tree in the CTA scan of the patient, and selection of surface landmark points nearby the coronary tree, 3) Estimation of coronary motion on the training datasets based on the selected cardiac landmark points, and 4) Creation of the coronary motion model. Note, that step 1 has to be performed once, while steps two to four have to be re-calculated for every new patient to be modeled. These proposed four steps are general, and could be implemented in many ways. We, for example, investigate two types of motion models for step 4: the classical statistical model defining the space of allowable motion estimates, and a shape conditional motion model using the cardiac shape of the patient to determine allowable motion estimates. The following sections describe our implementation of the coronary motion model step by step.

7.2.1 Cardiac training set with point correspondence

We automatically segment and landmark the four heart chambers and the aorta from a training set of 4D CTA scans, as described in [88]. End diastolic segmentation
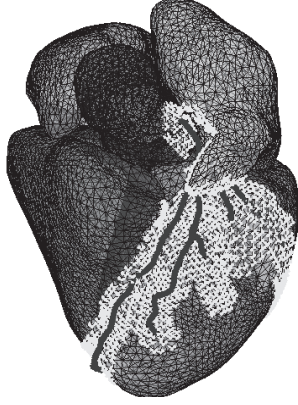
Figure 7.1 (color version on page 157): End diastolic heart segmentation (four chambers and the aorta), displayed together with the coronary arteries (blue) and the selected nearby landmark points (white).

is performed with multi-atlas registration using eight atlases. Point correspondence in the end-diastolic segmentations is achieved by defining corresponding landmark points on the atlases. The end-diastolic segmentations are then propagated to the other cardiac phases by 4D registration [89]. The resulting shapes are temporally aligned using the left ventricle volume curves to compensate for heart rate differences. Further details on landmarking, segmentation, and temporal alignment can be found in [88].

7.2.2   Coronary artery segmentation and landmark selection

The coronary artery centerlines of the patient to be modeled are extracted from the end-diastolic 3D CTA scan. We use a centerline extraction method based on a vesselness filter [46] and a minimum cost path between the aorta and a manually clicked vessel end point. This centerline is then refined using vessel wall segmentation based on [110]. A detailed description of the procedure can be found in [90].

The cardiac structures are segmented using multi-atlas segmentation as described in Section 7.2.1, resulting in surface points corresponding to the training set. Landmark points far from the coronary arteries are eliminated, and nearby landmark points are retained for further analysis. For each coronary artery centerline point, we select the closest 30 surface landmark points, resulting in a maximum distance of about one cm. Figure 7.1 shows an example segmentation with coronary centerlines in blue, and selected nearby landmark points marked white.

7.2.3 Coronary artery motion estimation on the training set

The end-diastolic phases of the training shapes are aligned based on the selected cardiac landmark points using translation and rotation. This alignment is propagated to the rest of the phases. One phase of a training shape can then be represented as a vector of its concatenated landmark coordinates $\boldsymbol{d_j}$, and the entire 4D training shape as the concatenation of all phases $j \in [1, T]$. The surface motion $\boldsymbol{l}$ is derived from this representation by subtracting the end-diastolic phase e from all phases[1]

$$\boldsymbol{l} = [\boldsymbol{l}_1, \boldsymbol{l}_2, \ldots \boldsymbol{l}_T] = [\boldsymbol{d}_1 - \boldsymbol{d}_e, \boldsymbol{d}_2 - \boldsymbol{d}_e, \ldots \boldsymbol{d}_T - \boldsymbol{d}_e] \quad . \tag{7.1}$$

The motion $\boldsymbol{m}$ of the artery centerline points is approximated for every training subject as the weighted average motion of their nearby cardiac landmark points k:

$$\boldsymbol{m}_j^i = \frac{\sum_k \nu_i^k \boldsymbol{l}_j^k}{\sum_k \nu_i^k} \quad , \tag{7.2}$$

where $\boldsymbol{m}_j^i$ denotes the estimated motion of *vessel* point i in cardiac phase j, and $\boldsymbol{l}_j^k$ denotes the motion of cardiac *landmark* point k in phase j. The weights ν_i^k give a higher weight to landmark points that are closer. We weighted landmarks with an exponential function of the distance $D(k, i)$, measured between landmark point k and vessel point i in the end-diastolic segmentation of the subject:

$$\nu_i^k = \exp(-D(k, i)/\sigma_v) \quad , \tag{7.3}$$

where σ_v controls the steepness of the weight function.

This step results in a motion estimate for every subject in the training set.

7.2.4 Statistical motion model (SMM)

To build a classical statistical motion model, principal component analysis (PCA) is used on the coronary artery motion estimates of the training data. The coronary motion for a given subject can then be approximated as

$$\boldsymbol{m} \approx \bar{\boldsymbol{m}} + \Phi \boldsymbol{b} \quad , \tag{7.4}$$

where $\boldsymbol{m} = [\boldsymbol{m}_1^1, \boldsymbol{m}_1^2, \ldots, \boldsymbol{m}_1^N, \ldots, \boldsymbol{m}_j^i, \ldots \boldsymbol{m}_T^N]$ contains the motion of all vessel points $i \in [1, N]$ in all phases j, $\bar{\boldsymbol{m}}$ is the mean motion, Φ represents the main modes of variation, and $\boldsymbol{b}$ is the vector of motion parameters. In this chapter we keep 95% variance in the model, assuming that the residual variance is caused by noise.

7.2.5 Shape conditional motion model

In previous work we showed that the motion of the four cardiac chambers and the aorta was correlated with their end-diastolic shape [88]. As we estimate the coronary

---

[1]Due to space considerations all vectors are represented as row vectors in Equation 7.1

motion from the cardiac motion, coronary motion is also expected to be correlated with the end-diastolic shape. The most probable motion $\hat{\boldsymbol{m}}$ of a coronary artery can therefore be estimated from the end-diastolic heart surface $\boldsymbol{d}_e$ nearby the artery via linear regression. In this work we use ridge regression, as that was found to slightly outperform other linear regression methods for large dimensional, small sample size data [8]. The regression can be formalized as

$$\boldsymbol{m} = \hat{\boldsymbol{m}} + \boldsymbol{\epsilon} = (\boldsymbol{d}_e - \bar{\boldsymbol{d}_e})\boldsymbol{\beta} + \bar{\boldsymbol{m}} + \boldsymbol{\epsilon} \tag{7.5}$$

where $\hat{\boldsymbol{m}}$ is the predicted motion, $\bar{\boldsymbol{m}}$ is the mean motion, $\bar{\boldsymbol{d}_e}$ and $\boldsymbol{d}_e$ are the population mean shape and the current shape of the end-diastolic heart surface nearby the coronary arteries, $\boldsymbol{\beta}$ is the vector of regression coefficients, and $\boldsymbol{\epsilon}$ is the remaining error. $\boldsymbol{\beta}$ is calculated from the training set as

$$\boldsymbol{\beta} = (\Sigma_{d_r} + \lambda I)^{-1} \Sigma_{d_r, m} \quad , \tag{7.6}$$

where Σ_{d_r} denotes the shape variance of the nearby region, $\Sigma_{d_r, m}$ denotes the cross-covariance between shape and motion, I is the identity matrix, and λ is the ridge parameter. We refer to [54] for a detailed description of ridge regression. The ridge parameter can be optimized by cross validation on the training set.

The predicted motion $\hat{\boldsymbol{m}}$ is the most probable motion for the coronary artery given the cardiac shape. To model the distribution of possible motions, we calculate the amount of motion variation from the remaining error $\boldsymbol{\epsilon}$ as

$$\Sigma_\epsilon = E\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\} =$$
$$\Sigma_m - 2\Sigma_{m,d_r} S^{-1} \Sigma_{m,d_r}^T + \Sigma_{m,d_r} S^{-1} \Sigma_{d_r} S^{-1} \Sigma_{m,d_r}^T \quad , \tag{7.7}$$

with

$$S = \Sigma_{d_r} + \lambda I \quad . \tag{7.8}$$

This equation can be derived by combining Equation 7.5 with 7.6, and substituting the resulting $\boldsymbol{\epsilon}$ into the left hand side of Equation 7.7. Similar to the SMM case, we can decompose the variance Σ_ϵ into its principal directions creating the conditional model

$$\boldsymbol{m} \approx \hat{\boldsymbol{m}} + \Phi_\epsilon \boldsymbol{b}_\epsilon \quad . \tag{7.9}$$

Also in this model, we keep 95 % of the variance of the training data.

The conditional model allows variations around the predicted motion rather than around the population mean motion and restricts the allowable motions more than the classical SMM.

7.3   2D+t/3D coronary centerline registration

A 2D-3D registration through time consists of two main components: the 2D-3D distance measure evaluating the alignment in one frame, and the strategy in which the frame-by-frame measures are combined and optimized to derive the alignment of the entire cardiac cycle. In this section we describe both components in detail.
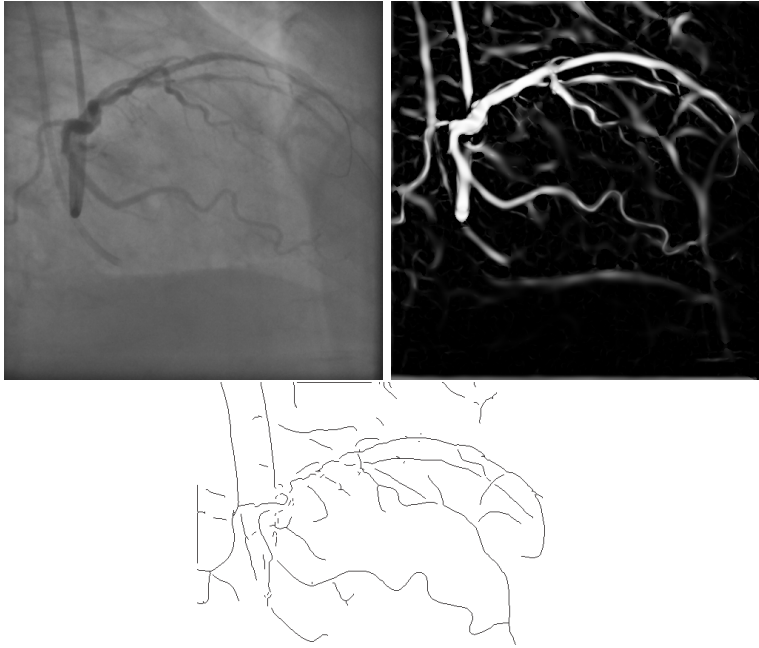
Figure 7.2: An example frame with from left to right: the original XA image, the vesselness filtered image, and the centerline image.

7.3.1 2D-3D distance measure

This measure quantifies the distance between the projected 3D vessel centerlines and automatically extracted 2D vessel centerlines in the XA image. The 3D vessels are projected to the image plane based on the geometry of the X-ray imaging system. The 2D vessel centerlines are extracted from the X-ray images by first enhancing dark longitudinal structures with a Frangi vesselness filter [46]. Subsequently, ridges are found via local non-maximum suppression, and hysteresis thresholding. Finally, single pixel long centerlines are removed. An example frame with original intensities, the vesselness filtered image, and the centerline image is shown in Fig. 7.2.

The 2D-3D distance measure of a frame is based on the distances between projected 3D vessel points and extracted 2D centerline points. To determine corresponding points and their distances, we adapted the distance function Θ_{ij} from [7], which comprises a 2D Euclidean distance and an orientation difference. We enriched this

measure with the Frangi vesselness score F_j at extracted centerline points j, such that

$$\Theta_{ij} = \alpha(1 - F_j) + (1 - |\cos\gamma_{ij}|\ e^{-\frac{D_{ij}^2}{\sigma^2}})\quad,\qquad(7.10)$$

where D_{ij} is the 2D Euclidean distance between projected vessel point i and centerline point j, and γ_{ij} is the angle between the 2D centerline direction and projected 3D vessel direction. The constant α regulates the importance of the vesselness term in the distance measure. Due to the exponential function all distances D_{ij} are mapped to the [0,1] interval, resulting in small gradients for large distances. The parameter σ regulates this non-linear behavior.

Two different methods to calculate frame-wise 2D-3D distances are proposed and used in the experiments.

For the first distance measure every 3D centerline point i is matched with a 2D centerline pixel j, such that Θ_{ij} is minimal. The final 2D-3D frame distance is the weighted average of the individual minimum distances:

$$G_1 = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \min_j \Theta_{ij}\quad,\qquad(7.11)$$

where N is the number of 3D centerline points, and weights w_i encode the distance of the vessel point i from its neighbors in the projection. This weighting ensures that all vessels have a similar influence on the optimization independent of their orientation with respect to the projection direction. Centerline points outside the FOV are weighted zero.

In the second measure, correspondences in both directions are created, meaning that also every 2D centerline point is matched with its closest projected 3D centerline point. The resulting symmetric frame distance measure is

$$G_2 = \frac{1}{\sum w_i} \sum_{i=1}^{N} w_i \min_j \Theta_{ij} + \frac{1}{M} \sum_{j=1}^{M} \min_i \Theta_{ij}\quad,\qquad(7.12)$$

where M is the number of extracted 2D centerline pixels.

7.3.2   2D+t/3D registration

We register a complete cardiac cycle by minimizing the sum of all frame-wise 2D-3D distances

$$G = \sum_{t=1}^{L} G_1(t)\quad\text{or}\quad G = \sum_{t=1}^{L} G_2(t)\quad,\qquad(7.13)$$

where L is the number of frames in the cardiac cycle. Minimization is performed by optimizing four groups of parameters: heart pose parameters relating the CT coordinate system with the angiography coordinate system, respiratory motion parameters, coronary artery motion parameters, and temporal alignment parameters. A more detailed explanation of the motion parameters and the temporal alignment is provided below. An overview of the number of parameters optimized in the different models is shown in Table 7.1.

Motion models		Number of Parameters			
Resp.	Coron.	Pose	Resp.	Temp.	Coron.
Quad	Fixed	6	12	1	
Quad	Statistical	6	12	1	K
Unconstr.	Fixed		6*L	1	
Unconstr.	Statistical		6*L	1	K

Table 7.1: The number of parameters optimized during 2D+t/3D+t registration with different motion model configurations. The number of frames of a cardiac cycle is denoted by L, and the number of modes of the statistical model by K.
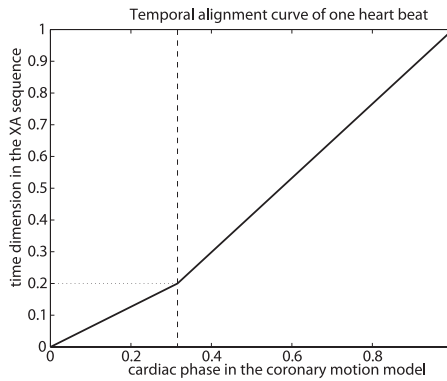
Figure 7.3: Temporal transformation function relating the end systolic time point in the motion model to a time-point in the XA sequence. In between phases and frames are calculated with linear interpolation.

Temporal alignment

The coronary artery motion model describes the shapes of the coronary arteries in different cardiac phases. A temporal alignment is needed to relate these cardiac phases to the XA frames, accounting for differences in temporal resolution and heart rate between the model and the angiography sequence. As heart rate influences the relative duration of systole and diastole, a piecewise linear function is proposed for the temporal alignment, with one parameter τ representing the relative length of systole in a cardiac cycle. The function is shown in Figure 7.3, with phases parameterized between 0 and 1.

Coronary motion models

The investigated coronary motion models can be grouped in two categories: motion estimates, and statistical motion models. While in the latter the motion pattern is optimized, in the first one the actual motion is fixed. Therefore, registration with a motion estimate is computationally cheaper.

The considered motion estimates are the 4D CTA motion as proposed in [90], the static end-diastole, where no cardiac motion is assumed, the mean motion of the training population, and the predicted motion from the cardiac shape as explained in Section 7.2.

The statistical models considered are the statistical motion model (SMM) and the conditional motion model as described in Section 7.2. These models were fitted by optimizing the motion parameters b in Equation 7.4 and 7.9 respectively.

Respiratory motion models

We investigate two strategies to account for the respiratory motion during XA acquisition. In both we consider only the rigid part of the respiratory induced movement of the heart, which was shown to account for the largest part of the motion [112, 113].

In the first strategy we model the respiration induced movement of the heart in one cardiac cycle as a quadratic function. In this case the first frame of the cardiac cycle is taken as reference with zero respiratory motion, and respiratory motion parameters are optimized for the center and the end of the cycle. Parameters include 3 degrees of freedom (DoF) for translation and 3 DoF for rotation, thus a total of 12 parameters are used to represent the respiratory induced motion in the entire cardiac cycle. The motion in any frame of the cycle is calculated from the start, middle and end time points using quadratic interpolation.

In the second, unconstrained strategy we optimize the pose of the heart in every frame solely based on the image information. Optimization of this approach is the slowest, due to the large amount of parameters (6 DoF per frame).

7.4   Implementation details

7.4.1   Optimization

We used the non-linear least squares optimizer implemented in the optimization toolbox of Matlab R2011. This is a subspace trust-region method based on the interior-reflective Newton method. We set 20 mm and 20° bounds on the initial displacement parameter, and 10 mm and 10 degree bounds on the respiratory model following [113]. These bounds ensure that the out of plane motion stays within realistic bounds. The motion model parameters were restricted to lie within 3 standard deviations from the mean shape for both the classical statistical and the conditional models. In order to increase registration speed we sub-sampled the automatic 2D centerlines by a factor of 6 resulting in a neighboring centerline point distance of about 1.5 mm, and the 3D coronary artery centerline by a factor of 3 resulting in a 3D centerline point distance of about 2.4 mm.

Our optimization strategy was slightly different for the different breathing models. When using the quadratic breathing model, all parameters were optimized at once (parameters were weighted such that a unit change caused the same order of magnitude difference in the distance measure as a unit translation in plane). When optimizing the unconstrained breathing we alternately optimized the respiratory motion

parameters, and shape parameters (including the temporal fit) to keep the number of parameters to optimize within reasonable limits.

7.4.2 Parameter settings

The vesselness filter parameters used for the automatic centerline extraction were set as proposed in [46]. Vesselness was calculated on three scales (0.5, 1, and 2 mm). The high and low threshold values for the hysteresis thresholding of the vessel centerlines were manually set to 0.45 and 0.1 after visual inspection of 3 sequences.

For the coronary artery motion estimation from the cardiac surface motion we set the parameter σ_v in Equation 7.3 to 4 mm. In this manner the furthest selected landmark points are weighted less than 0.1 in the motion estimation.

The σ^2 parameter of Equation 7.10 was set to 50 mm^2, preventing distances larger than 15 mm to influence the alignment. The weighting constant α of the same equation was set to 0.1.

7.5 Data

7.5.1 Training data for the motion model

The 4D heart segmentations from [88] were used to build the statistical motion models of the coronary arteries. The data consisted of 151 ECG gated 4D CTA images acquired as part of the clinical diagnosis of patients with acute or stable chest pain. The data therefore includes a large variety of anatomies and pathologies. CTA reconstructions were made every 5% of the cardiac cycle resulting in 20 images per sequence. The voxel size of the 3D+T CTA sequence was approximately 0.7x0.7x0.9 mm^3.

7.5.2 Test data

For evaluation, 12 XA datasets of 10 patients (two patients underwent two interventions) were used. The same patients were analyzed in [90] for the case of known patient specific coronary motion. Images were acquired between 2007 and 2009 using a Siemens Axiom Artis biplane system. As no calibration data was available, in this work only sequences of the primary C-arm were included. Images contain 512x512 pixels with a pixel size of 0.22x0.22 mm^2. We selected two consecutive cardiac cycles from each sequence for analysis, both cycles containing contrast enhanced coronary arteries. The average number of frames in a cardiac cycle was 13 frames, and ranged from 9 to 22 frames. One sequence was acquired at a frame rate of 30 frames per second (fps), the others at 15 fps. Five of the sequences contained the right coronary artery, the others the left coronary tree. Every patient in the test set underwent a 4D CTA scan prior to PCI. These CTA scans were also part of the training data described in the previous subsection. To derive the patient specific coronary artery motion from the 4D CTA, we first extracted the centerlines at end-diastole as de-
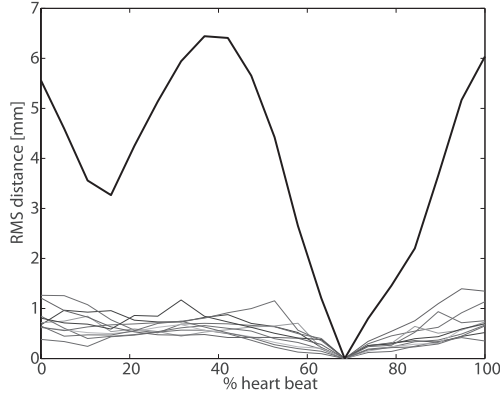
Figure 7.4: Evaluation of the estimated coronary motion from the nearby cardiac surface. The colored time-curves show the rms distance of the extrapolated coronary arteries to the extracted arteries from 4D CTA for all 10 patients. The bold black line serves as a reference, and shows the average rms distance of the coronary arteries from their end-diastolic phase.

scribed in Section 7.2.2, and then propagated these to the other cardiac phases. We refer to [90] for further details on 4D centerline extraction.

7.6   Experiments and results

We performed two sets of experiments. The first set of experiments focused entirely on the coronary motion model, which was evaluated independently of the task of 2D+t/3D registration. The second set of experiments focused on the 2D+t/3D registration and assessed registration accuracy as well as the quality of the final motion estimate.

7.6.1   Coronary motion model evaluation

In these experiments the coronary artery locations extracted from the 4D CTA datasets of the 10 test patients were used as ground truth.

Coronary motion extrapolation from nearby heart landmarks

We evaluated the accuracy of estimating coronary motion from the motion of nearby cardiac surface landmark points. The resulting coronary artery centerlines were compared to the arteries extracted from the 4D CTA.

Results indicate an average root mean square (RMS) point-to-point distance of 0.53 mm. Figure 7.4 shows the distance over time for all 10 patients. As a reference the average coronary motion (mixed left and right trees) is plotted in black.

Table 7.2: Generalization evaluation of the motion model. We report the average and standard deviation of the 3D point-to-point errors of all phases.

	Accuracy [mm]
SMM representation	1.02 ± 0.29
Conditional model representation	1.02 ± 0.29
Predicted motion	2.29 ± 0.43
Mean motion	2.52 ± 0.74
Static	4.4 ± 1.39

Model generalization

The generalization power of a statistical model indicates the accuracy with which the model is able to represent a previously unseen subject. We evaluated generalization on the 10 test subjects, with a model built on 150 training samples (excluding the current test subject). The statistical models contained about 45 modes of variation after dimensionality reduction, and the parameters of each mode were restricted to lie within the ± 3 standard deviation bound. The representation accuracy of the model was compared with the mean motion, predicted motion, and static estimates. Prediction was performed as described in Section 7.3.2, with λ, the ridge parameter being optimized on the training set excluding the test subjects.

The results in Table 7.2 indicate that both statistical models were able to represent the motion with about 1 mm average root mean square point-to-point accuracy, and thereby significantly outperformed the motion estimates. The predicted motion tends to have a better accuracy than mean motion, though the difference was not found to be significant. The static motion estimate, as expected, performed significantly the worst. All significance tests were performed at the 5% significance level with a paired t-test.

7.6.2 2D+t/3D registration evaluation

We compared in total 12 registration strategies varying both the coronary and breathing motion models respectively. Coronary motion models were built on 150 training subjects leaving the test subject out from the training set. Every registration strategy was run nine times on all test data with different manual initializations. Three manual initializations were collected from three independent observers, who were asked to translate the end-diastolic coronary artery in the projection plane (2 DoF) such that the projection of the 3D vessels matched the first angiographic frame of the cardiac cycle well.

The XA data contained images of patients with metallic stitches from previous heart surgeries, and in some cases the guiding catheter was visible in the field of view (FOV). We therefore performed stitch and catheter removal in a preprocessing step. We first semi-automatically delineated the stitches in the first frame using a live-wire approach, and then tracked their position using intensity based rigid registration

through time. Finally the stitches were masked, and replaced with the gray-value morphological closed image intensities. A similar strategy was used for the catheters.

Corresponding centerline distance For evaluation purposes, we adapted the centerline distance measure from [90]. First, centerline segments were manually drawn on the XA frames. The corresponding closest projected 3D centerline branch for every annotated centerline segment was determined, and subsequently the distances of all points in the annotated centerline segment to the projected 3D centerline branch were calculated. The root mean square (RMS) distance of all annotated centerline points resulted in the corresponding centerline distance measure for each frame.

Experiment 1: Comparison of motion models

We compared registration performance of the coronary and respiratory motion models described in Section 7.3.2. One cardiac cycle per sequence was used for evaluation, with nine different manual initializations each. We performed registrations with both 2D-3D distance measures, G_1 and G_2. In prior experiments we found that due to the numerous false 2D centerline detections, the convergence range of G_2 was smaller than that of G_1. Therefore, in our experiments G_2 was always applied after an optimization with G_1.

Table 7.3 shows the median centerline distances of all frames with both distance measures. [2] To check for statistical significance, we averaged all results per patient, and applied a paired Wilcoxon signed rank test on these averages. Significant differences between the proposed methods and the 4D CTA motion based fitting are reported at the 5% significance level in Table 7.3 as bold typeset. All results were significantly different from the initialization. Initializations using the mean motion gave a median centerline distance of 3.88 mm. Figure 7.5 shows the centerline distances achieved with the G_2 distance measure. The numbers at the top indicate the number of frames with a centerline distance larger than 6 mm. Example alignments for one frame with different coronary motion models and quadratic breathing are shown in Figure 7.8. Figure 7.9 shows further results for three time-points of four patients using the 4D CTA motion and using the statistical coronary motion models, with the quadratic breathing model.

We investigated the rigid cardiac motion obtained with the unconstrained breathing optimization, and plotted the resulting pose parameters in the in- and out- of plane directions. An example is given in Figure 7.6. We can make the following observations: 1) The out of plane pose has a large variation, 2) the motion pattern of the static estimate considerably differs from the other motion models, 3) the time-curves of the population based motion estimates and statistical models are close to each other, and 4) The 4D CTA motion curve is in some cases close to the population models, in some cases not.

We also evaluated the difference between the estimated motion of the statistical models and the 4D CTA motion, while focusing strictly on the intrinsic cardiac

---

[2]We report results on all 10 available datasets, as the registration accuracy on the datasets used for setting the method parameters were not better than the results of the other datasets.
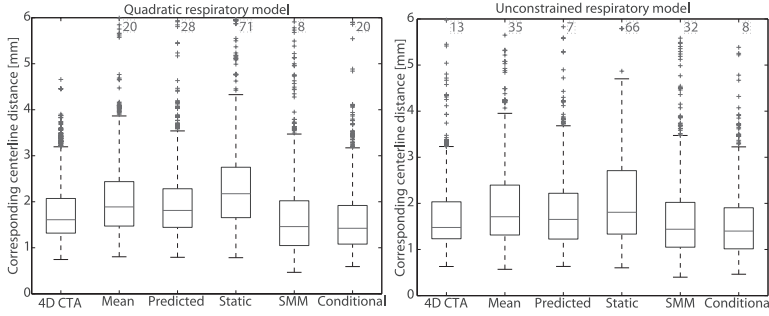
Figure 7.5: Boxplots of the 2D corresponding centerline distance measures of all frames and all patients with the G_2 distance metric. From left to right: quadratic respiratory motion model, and unrestricted respiratory motion. The numbers at the top show the number of outliers outside the range of the current plot.

Table 7.3: Median 2D corresponding centerline distance measures [mm] of all frames in one cardiac cycle of all patients. The first row indicates the used coronary motion model, and the first column the used breathing model (quadratic and unconstrained). Bold numbers represent statistically significant differences with the Wilcoxon signed rank test at the 5% level, compared to the 4DCT motion fit of the same row.

	4DCT	Mean	Pred.	Static	SMM	Cond.
G_1						
Quadr.	1.82	**2.15**	2.09	**2.65**	1.74	1.69
Unc.	1.83	**2.17**	1.94	**2.54**	1.89	**1.67**
G_2						
Quadr.	1.61	**1.89**	1.81	**2.17**	1.46	1.42
Unc.	1.48	**1.71**	1.65	**1.81**	1.44	**1.40**

motion. We separated this motion by reconstructing it from the optimized coronary artery motion model parameters. To gain insight in the directional dependency of the error, we report in-plane and out-of-plane 3D motion differences. A box plot of the per patient per initialization average of these errors is given in Figure 7.7, showing no obvious direction dependency.

Experiment 2: Reusability of the optimized motion

In this experiment we investigated whether the resulting coronary artery motion estimate of one cardiac cycle can be applied for the 2D+t/3D+t registration of a subsequent cycle. Hereto, we fixed the resulting motion from the first cardiac cycle, and only optimized the pose and respiratory motion for the next cardiac cycle. We compared the accuracy of this approach with the full optimization of heart pose, respiratory-, coronary-, and temporal parameters. The heart pose was initialized by

Figure 7.6 (color version on page 158): In plane (left, middle) and out of plane (right) translation for patient 2 through time when using the unconstrained respiratory motion model. For the color image see page 158.
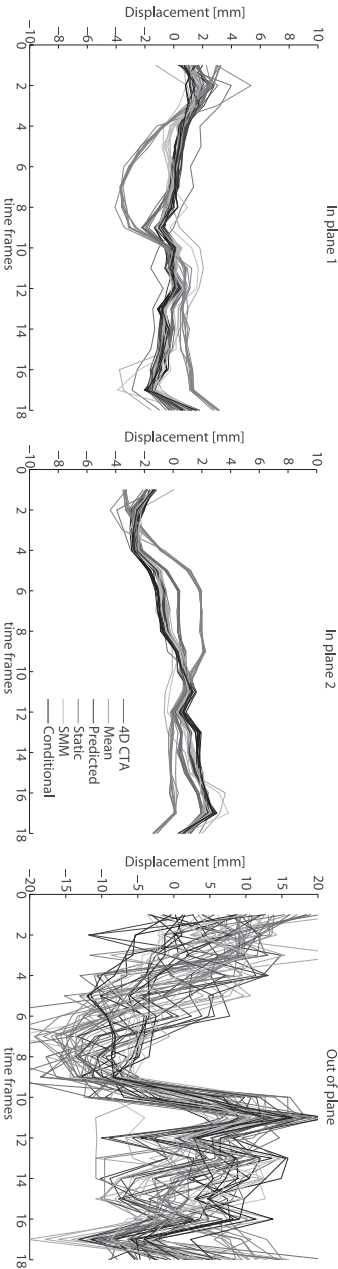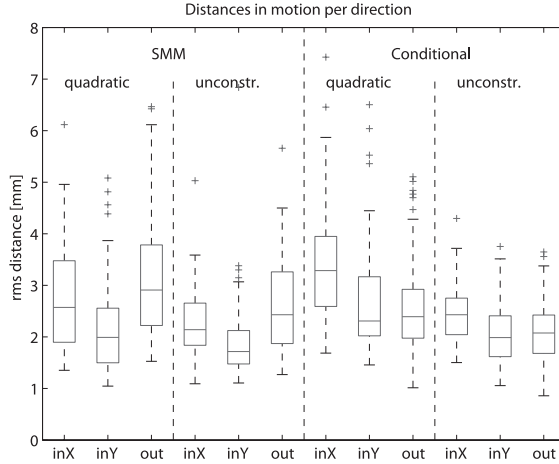
Figure 7.7: RMS point-to-point errors of the optimized coronary motion (SMM), and the conditional coronary motion compared to 4D CTA motion. Columns are separated into in-plane and out of plane directions. Both quadratic and unconstrained respiratory models are shown.

Table 7.4: Reusability of the optimized coronary motion with quadratic respiratory model. The table shows the median 2D centerline distance [mm] of all frames of all patients in the second cardiac cycle with initialization based on the first cardiac cycle. In the first row the coronary motion estimate from the first cycle was used, in the second row all model parameters were newly optimized. Bold numbers represent statistically significant differences from the 4DCT at the 5% level within a row.

Parameters optimized	SMM	Cond.	4DCT
Pose and respiration	1.74	**1.78**	1.90
All	1.83	1.81	1.88

extrapolating the quadratic breathing motion of the first cardiac cycle. Registration was performed with the G_2 distance measure.

Table 7.4 summarizes the results, and shows that the reused motion (first row) performs similar to, or slightly better than, a complete new fit of the cardiac motion model (second row).

7.7 Discussion and Conclusions

In this chapter we investigated the possibility of using coronary motion models combined with a single end-diastolic CTA acquisition for aligning the CTA data with dynamic XA data. Several 4D coronary motion models were investigated, including motion estimates representing a specific motion pattern (such as the mean motion of
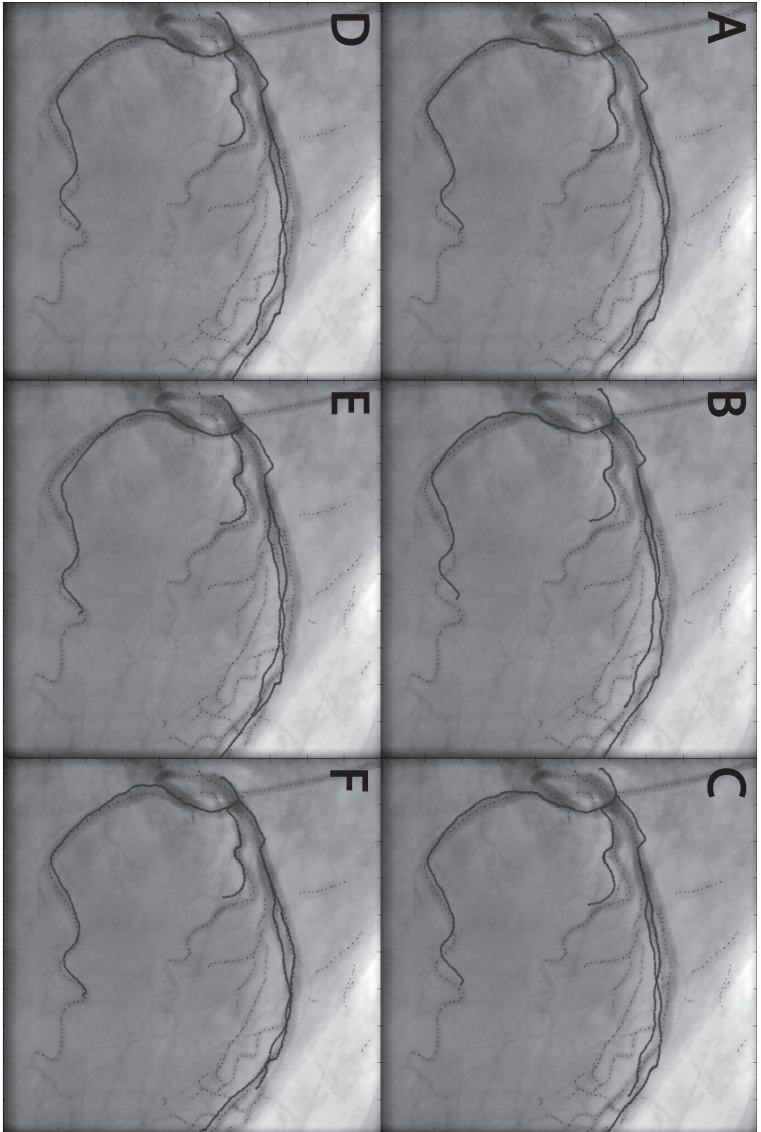
Figure 7.8 (color version on page 159): Example alignments with all coronary motion models and the quadratic respiratory model. The blue lines are the fitted coronary arteries overlayed on the corresponding systolic X-ray frame. Black dots show the subsampled centerlines used for fitting. A) 4D CTA motion, B) mean motion, C) predicted motion, D) static end-diastolic estimate, E) statistical motion model, F) conditional motion model.

a population, the predicted motion and no motion), and statistical models, which are able to adapt the motion pattern to the image information.

The proposed statistical models were constructed by extrapolating cardiac surface motion close to the coronary arteries. This extrapolation resulted in an average root mean square point-to-point accuracy of about half a mm, which is smaller than the voxel size of the CTA scans (Section 7.6.1 and Figure 7.4).

Both the classical statistical motion model and the conditional statistical motion model were able to represent previously unseen motions with an average accuracy of about one mm (Table 7.2). The mean of the conditional model tended to predict the coronary motion better than the population average. However, most probably due to the small sample size, this difference was not found to be statistically significant.

We compared the 2D+t/3D registration accuracy of the coronary motion models and estimates using two different distance metrics, G_1 and G_2 (Table 7.3). The trends obtained with both metrics were the same. The 4D CTA motion consistently outperformed the mean and predicted motion estimates. The best results, however, were achieved by the statistical motion models. This might be due to slight differences in the cardiac cycle between intervention time and CTA acquisition. Such differences can be caused by e.g. respiratory induced non-rigid cardiac motion, slight variations in the cardiac motion pattern, or small inaccuracies in the selection of the start and end frames of a cardiac cycle in the XA sequence. The statistical motion models have the advantage to be able to adapt to such changes in the cardiac cycle, and hence outperform the reference standard motion derived from a 4D CTA acquisition. The conditional and classical statistical motion models did not show a large difference in performance, though the conditional model tended to have somewhat lower median centerline errors. These trends hold for both quadratic and unconstrained respiratory motion models.

The additional degrees of freedom in the unconstrained respiratory motion model enabled compensating some of the shape variations during a cardiac cycle. This is especially visible with the static end-diastolic coronary motion model, where the difference in accuracy using the quadratic and unconstrained respiratory models was the largest (Table 7.3). The same conclusions can be drawn from the inspection of the time-curves of the resulting pose parameters of the unrestricted respiratory motion models (Figure 7.6). The static estimate typically had a larger respiratory motion, and a different respiratory pattern than the other motion models.

Though the employed 2D evaluation metric shows better results with the unconstrained respiratory model than with the quadratic model (Table 7.3), the resulting out of plane motion with the unconstrained respiration seems implausible. This might be due to the small appearance changes in the projected centerline when moving toward or away from the source, and could also be an artifact of the down-sampling applied to the automatic 2D centerlines. When investigating the effect of the projection direction on the coronary motion pattern, no obvious direction dependency was found (Figure 7.7). This might be caused by the global nature of the motion modes resulting from the PCA analysis, equally affecting all three dimensions.

We furthermore found that using the estimated coronary motion of one cardiac cycle in a consecutive cycle, performed similar to, if not better than a complete new

registration with the statistical models (Table 7.4). A possible explanation is the lower contrast in the vessels due to wash-out, which caused some of the centerlines detected in the first cycle to be missed in the second cycle. The derived motion from the first heart cycle also outperformed the registration using the 4D CTA derived motion on the second cycle. This result indicates that the motion parameters, though highly varying for registration of the same patient with different initializations, hold meaningful information about the current patient specific contraction pattern.

As no geometrically calibrated second view was available in our dataset, we were not able to evaluate the correctness of the estimated cardiac motions in 3D. Therefore, in this chapter, we solely rely on accuracy measurements on the projection plane. The measured overlay quality is important for several applications, such as for over-projecting lesion information from CTA on the XA sequence.

Registration time required for the evaluated methods depend on the number of frames in the cycle, and the number of parameters optimized. While the final goal would be a real time registration, our unoptimized Matlab implementation takes about an hour to run on a current laptop computer for an average length cardiac cycle using a statistical coronary motion model and quadratic breathing. We believe this time can be greatly reduced by optimizing the code and parallelizing computations.

In conclusion, this chapter proposed the construction of population based coronary motion models, and showed their applicability for turning the 2D+t/3D registration of end-diastolic CTA and XA sequences into a 2D+t/3D+t registration. The registration performance using the motion models was similar, if not better than the registration performance using the patient specific motion extracted from 4D CTA. This result was found for all tested distance measures and respiratory models. Furthermore, the fitted coronary motion patterns were applicable for registration of a subsequent cardiac cycle with lower contrast. Therefore, we believe that statistical models may substitute the patient specific motion for the task of 2D+t/3D+t alignment of XA sequences and patient specific coronary arteries.

7.8   Acknowledgements
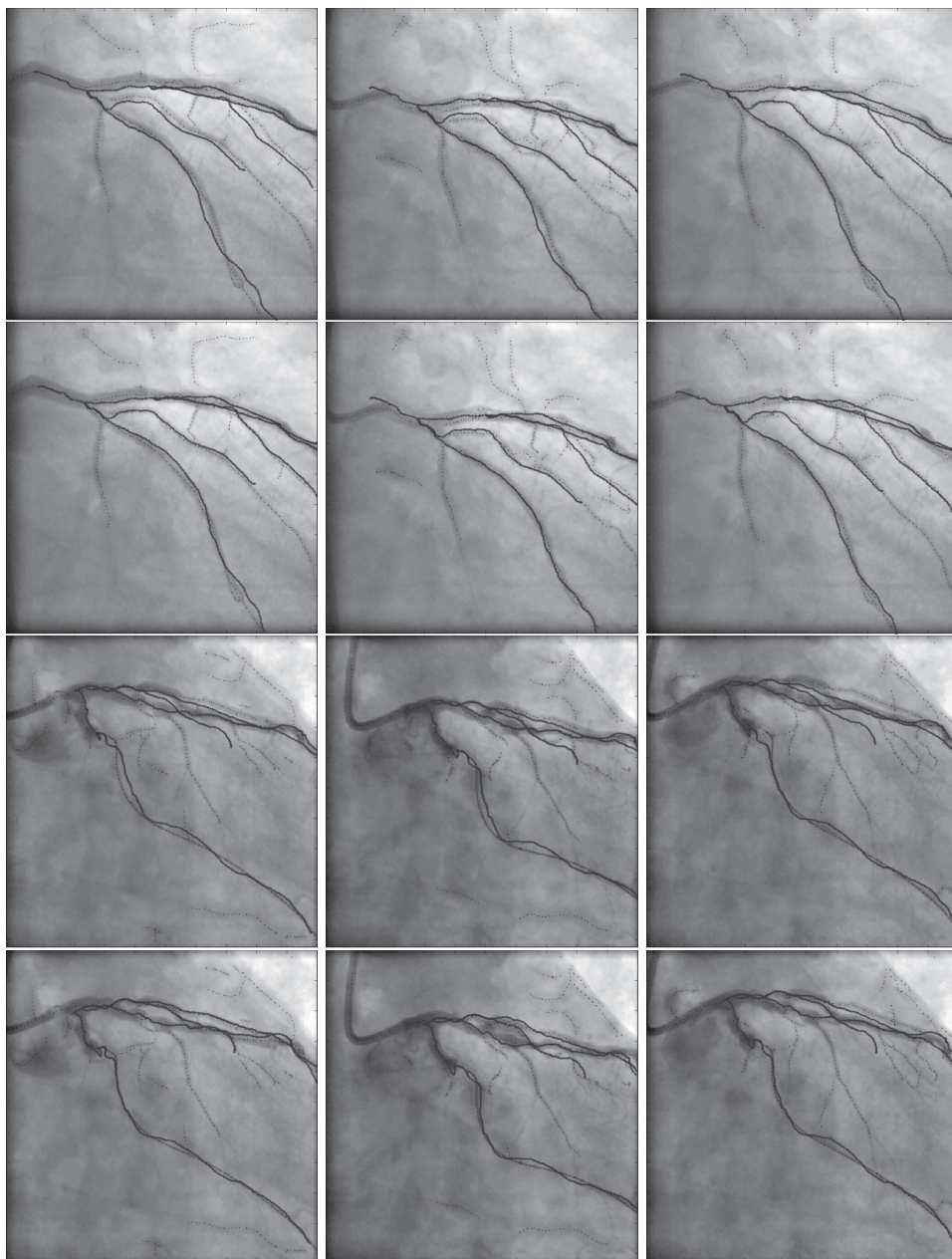
Figure 7.9 (color version on page 160): 2D+t/3D registrations of 3D vessels (blue) to XA data of two patients. Images are taken at 15%, 50%, and 85% of the cardiac cycle. First and third rows use the 4D CTA based coronary motion, the second and last rows use the SMM. A quadratic respiratory model was used in all cases.

Figure 7.10 (color version on page 161): 2D+t/3D registration results (continued). Images are taken at 15%, 50%, and 85% of the cardiac cycle. First and third rows use the 4D CTA based coronary motion, the second and last rows use the SMM. A quadratic respiratory model was used in all cases.
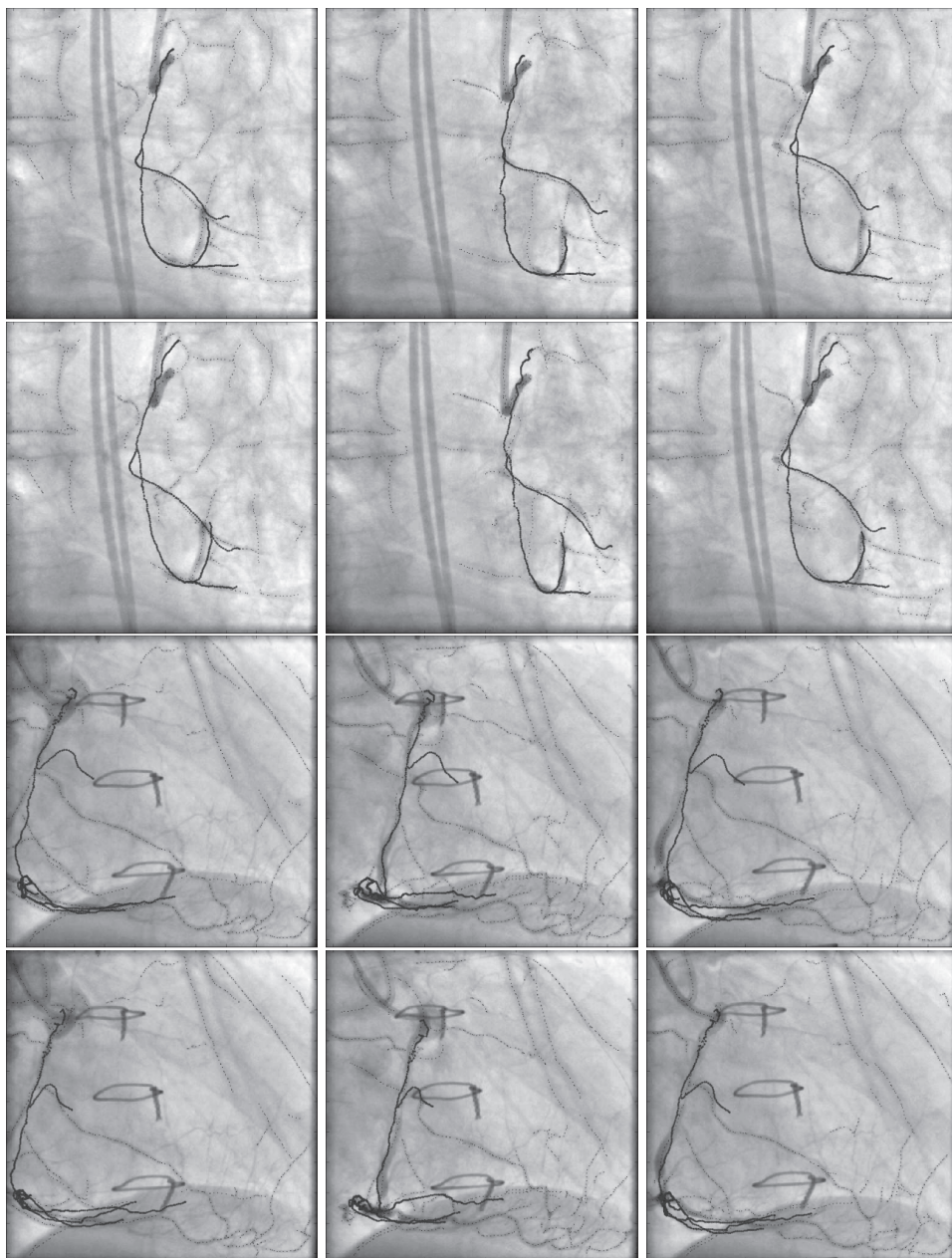
Chapter 8

Summary and Discussion

8.1   Summary

In this thesis we investigated several sparse image acquisition scenarios, and applied statistical models to complement the sparse data. The focus of this work was therefore not on solving one specific medical problem, but rather on investigating the applicability of a method to a wide range of problems. In the course of the work several medical applications were addressed, being bone reconstruction for orthopedic surgeries and for injury prevention, and heart motion estimation for percutaneous coronary interventions.

The first chapters entirely focused on bone reconstruction from bi-plane X-ray fluoroscopic images. In **Chapter 2** we proposed a method that reconstructs the patient specific femur from partial projections. The developed method is robust to non-orthogonal acquisition setups, which are standardly employed e.g. for Roentgen stereophotogrammetric analysis (RSA). Also, the method minimizes user interaction to a starting pose for optimization, for which we enlarged the convergence interval with a simple annealing technique. Moreover, due to an angle weighting the method is robust against parallel edges with opposite gradients, as often occurs at joints. We achieved a 1.68 mm average root mean square point to surface accuracy for the distal femur, and have shown that reconstruction of the entire femur from distal projections was possible. The accuracy of the proximal side is though expected to be lower than that of the distal part [9].

In **Chapter 3** the focus shifted from the general 2D-3D reconstruction to the more specific problem of biplane kinematic femur tracking and reconstruction. In the kinematic tracking an extra temporal dimension was added to the 2D-3D reconstruction of Chapter 2. We changed our method accordingly, and incorporated all frames in the shape estimation. This was shown to outperform in both accuracy and robustness the single time-frame reconstruction approach. Furthermore, to keep the user interaction minimal, i.e. a single manual pose initialization per sequence,

we proposed to integrate appearance priors in the edge selection. These appearance priors improved the convergence rate from 71% to 91%. The achieved average reconstruction accuracy of the distal femur was 1.48 mm root mean square point to surface distance, and the tracking precision was about 1-1.5 mm translation and 1-1.5° rotation. The most difficult parameter to optimize was the rotation around the longitudinal axis of the femur, which had an accuracy of 2.5° . This is most probably due to the tubular shape of the distal femur causing only relatively small differences in the projected silhouette under rotation.

While in Chapters 2 and 3 the point correspondences between the image edges and model were not known a-priori, in the following chapters we assumed known point correspondence. This can be achieved by correspondence estimation [141], or can be known by the sparse data acquisition [92, 88]. What is especially interesting in this case, is that the shape estimation problem can be solved analytically rather than by numerical optimization. When solving the shape estimation problem analytically, we aligned the training set for model building on the same landmarks as were available for the personalized shape estimation. This ensured that with solely one pose estimation the model was able to generate all training shapes.

In **Chapter 4** we used principal component regression (PCR) for deriving the cardiac motion of a subject based on the shape of the cardiac structures at end diastole. We hypothesized that the shape and motion are not independent, and show that the cardiac cycle estimated on the basis of the end diastolic cardiac shape has an average point to surface distance of 1 mm. In a follow-up work Metz et al. [88] showed, that the cardiac motion estimated with PCR slightly but significantly outperformed the population average cardiac motion.

The end diastolic heart surfaces and landmarks in Chapter 4 were computed using multi-atlas segmentation, meaning that several CTA images were registered using B-spline registration to the end diastolic patient CTA. The segmentations of these registered CTA images could then be propagated to the patient image, and could be averaged to derive the final segmentation. The spread of the separate segmentations around the average indicated the uncertainty of the segmentation at every landmark point. This example shows clearly, that image derived features such as points and surfaces are rarely, if ever, exact. Image quality and lack of distinctive features are possible causes for uncertainty. In **Chapter 5** we therefore introduced a new approach for regression based shape reconstruction, which takes the uncertainty of the known data into account. This maximum a-posteriori ridge regression (MAPRidge) was evaluated for both prediction accuracy, and resulting confidence interval on 2D heart MRI cross sections with simulated uncertainty patterns. Results showed a slight improvement in performance with the proposed method compared to the conventional ridge regression.

Estimating the point uncertainty, however, is not always trivial. Therefore, in **Chapter 6** we investigated in which situations it is beneficial to incorporate this additional knowledge in the shape estimation. We employed two methods that include point uncertainty (the MAPRidge regression as proposed in Chapter 4, and maximum likelihood principal component regression), and compared them to standard regression methods often used in literature (ridge regression, principal component

regression and partial least squares). Many regression methods have been proposed for shape regression up to date [3, 16, 34, 69, 77, 106, 133], without consensus about which method works the best in this scenario. We performed experiments on simulated data as well as real world models, and grouped the uncertainty scenarios in two categories. In one category the uncertainty of the test data was similar to the training data, in the second group the uncertainties differed largely. We found that in situations where the test data had an uncertainty pattern considerably different from the training data, incorporating point uncertainties using the MAP ridge regression provided the best estimates. However, when both the training and test data had similar uncertainties, the difference between standard methods and methods taking the uncertainty into account vanished. Moreover, in all our experiments we found that ridge regression slightly, but significantly outperformed the other standard regression techniques.

Chapter 7 rounds up the thesis by connecting the 2D-3D reconstruction work presented in chapters 2 and 3 with the cardiac motion estimation work presented in chapters 4 and 6. In this chapter we proposed a methodology for constructing statistical motion models of the coronary arteries based on CTA images. As the coronary topology is highly varying, direct modeling would be cumbersome. Hence, we proposed to derive the coronary artery motion from the motion of the near-by cardiac surface. In this way we were able to approximate coronary motion on a training set of segmented 4D CTA images, and subsequently construct its statistical model. Besides construction of the standard statistical motion model, we also constructed a model conditioned on the end-diastolic cardiac shape. We performed ridge regression for the conditioning, as that was shown to perform best in Chapter 6. Conform to Chapter 4 and its extension [88], the predicted motion tended to outperform the population mean motion in 3D-3D point-to-point accuracy. The statistical motion model parameters were optimized by minimizing a feature based 2D/3D distance metric. Similar to Chapters 2 and 3, the distance metric contained both Euclidean and angular distance. The features employed in this chapter, however, were not the edges in the image as in the first chapters, but the possible centerlines obtained by thresholding the ridges of the vesselness filtered image. Results showed that the 2D+t/3D registration with the statistical motion models performed at least as good as the 4D CTA based motion estimate. Moreover, the conditional motion model had a slightly better median accuracy than the classical motion model. We believe that these results are generalizable, as the two tested distance metrics and the two different respiratory motion models all resulted in the same trends.

8.2 Discussion

8.2.1 Methodology and possible extensions

In this work we presented statistical shape model based dense shape reconstruction from sparse data. Our main focus was on 2D/3D reconstruction of rigid and deforming shapes from still images and time sequences, and on shape prediction from sparse point data, such as 3D/4D prediction. In all cases we employed statistical knowledge

for the reconstruction, either in a form of a statistical shape model, or directly from the training set distribution. Our results showed that the presented approaches are viable, and able to produce a good estimate of shape and motion.

Shape estimates from sparse data, however, are never perfect. Part of the errors are due to the ambiguity of the problem, as multiple shapes may perfectly match the sparse data. The best possible shape estimate is therefore the shape that matches the sparse data and is the most probable shape based on the training set. The confidence interval of this shape estimate shows the remaining variability after matching the sparse data. We investigated and calculated confidence intervals for shape prediction in Chapter 5 and in Chapter 7, where the remaining variability of the predicted motion was used to constrain the refinement of the motion based on the XA data. The proposed approach to calculate the reconstructed shape confidence integrates the pose variation with the shape variation by aligning all training shapes based on the known model points. This enables more realistic estimation of the confidence than only modeling shape variation. This work could be further extended to deal with unknown correspondences, which occurs in many real situations. The extended problem, however, cannot be solved analytically. Recently, non-parametric estimation of the reliability has been proposed for the case of unknown correpondences [17].

Another source of error in statistical shape reconstruction is associated with the quality of the shape model. In practice, training sets sizes are in the same order of magnitude as the intrinsic shape dimensionality, consisting of tens or rarely hundreds of training shapes. As a consequence, the estimated distributions are imperfect, and the model is not able to accurately generate all shape instances. This could limit practical applications where accuracy is crucial. However, several ways to overcome this limitation have been proposed in the literature. The variation included in the model could be artificially enlarged by either directly adding variances to the covariance matrix [29, 35, 81], or by adding plausibly deformed shapes to the training set [81]. Also, the shape instantiation could be improved by a post-processing step allowing for small deformations to perfectly fit the known data. In the case shape model and image information are of the same dimensionality, dynamic programming was shown to improve segmentation [76, 98]. In the sparse case, however, further constraints need to be introduced, such as smoothness of the deformation field [48, 138, 140, 141]. This latter approach is generally applicable, and combines the strength of both statistical shape models constraining the registration, and non-rigid deformations allowing to leave the shape space and match the underlying data more precisely. Extending our presented methods with any of these techniques, and/or considerably enlarging the training set of the shape model is expected to increase the achieved accuracies.

Yet another source of error in statistical shape reconstruction is due to the uncertainty, or error in the sparse data. We dealt with this uncertainty in Chapter 5 and 6 for the case correspondences between sparse data and model are known. We have shown that the knowledge of the uncertainty of the data may improve reconstruction accuracy, if the training set uncertainty is considerably different from the sparse data uncertainty.

Part of this work focused on 2D/3D shape reconstruction and tracking. We ap-

plied these methods for femur reconstruction as well as for coronary artery tracking. These two anatomies differ largely, but we showed that a similar edge/centerline based registration works well in both domains. We chose the feature based registration for its favorable computational properties. Another option could have been intensity based registration using digitally reconstructed radiographs (DRR). DRR based methods are gaining popularity for *rigid* 2D/3D registration due to fast hardware acceleration techniques. A comparison study suggested, that DRR based rigid registrations outperform feature based methods in the neuro-intervention setting [86]. For non-rigid registration such comparison has to date not been performed. In the non-rigid case an appearance model is used for reconstruction, including statistical information on shape and intensities [132, 137]. As the metric optimized is based on the intensity differences between model generated intensities and X-ray data, such registration resembles the active appearance model (AAM) fitting in the dense imaging case. Comparison studies on medical and on non-medical images suggested that AAMs do often not outperform the simpler edge based active shape models (ASM) for segmentation accuracy [28, 130], though in some situations AAM may outperform ASM [123]. It hence remains unclear if DRR or edge based non-rigid 2D/3D reconstruction techniques perform better, and in which situations. We believe, that intensities do hold extra information that can make reconstruction more robust than a simple edge based approach, but (inner and outer) silhouette edges constitute the most relevant part of the image information. Therefore, in Chapter 3 we introduced a novel method that integrates advantages of both the edge based methods and the intensity based methods by combining the sparse edge representation and the appearance information at edge locations into one sparse framework. This was shown to outperform simple edge based non-rigid registration. Currently, the type of features calculated from the image intensities and the feature parameters are set manually. An interesting extension to our work would be to automatically extract the relevant features from training images. This could be achieved using machine learning methods, treating the manually selected edges and the X-ray images as the training set.

Motion tracking from fluoroscopic sequences, as proposed in Chapter 3, is of great interest for kinematic research, for surgical outcome assessment, and injury diagnosis. In most of these cases the relative motion of bones in a joint is of more interest than the motion of a single bone. A straight forward extension of our work is therefore the incorporation of the tibia in the reconstruction and tracking. Our methodology is general enough to make this inclusion possible with no, or little expected changes to the method. However, inclusion of the tibia also offers new possibilities to increase the robustness of the method, by introducing neighborhood constraints in the matching e.g. by penalizing too large surface distances or intersections between the surfaces. It has also been shown that bone surfaces at joints are highly correlated [34, 133]. This correlation could be exploited for a more robust shape estimation in case a sufficiently large training set is available for model building. Furthermore, the reconstruction and tracking accuracy of the method may be improved by slightly modifying the 2D-3D distance function employed. Our experiments in Chapter 7 showed that the sym-

metric rather than the one-way distance metric substantially improved registration accuracy. Symmetric distances improve point correspondence when no outliers are present, but are easily deteriorated by outliers such as noise edges. Therefore, application of a symmetric distance metric after the rough alignment is expected to improve tracking accuracy. Yet another interesting extension of our work could be to enforce temporal smoothness of the motion, for example using a Kalman filter. We omitted this constraint in our current work, as one patient showed abrupt changes in its motion curve. Therefore, before employing such a prior, further investigation is needed to identify the cause of the observed abrupt change.

The second part of the thesis focused on modeling the motion of the heart, and the coronary arteries from 4D CTA images. During cardiac systole, the heart muscles contract, and perform a small rotation movement. This rotation is not contained in our training set, as movements along the organ boundary cannot be extracted from CT images. A more accurate, though coarser resolution myocardial motion can be extracted from tagged MRI sequences. The combination of motion extracted from CTA and tagged MRI could therefore improve the motion estimate, for sure in the left ventricle. The motion models in this thesis were constructed on the surface points of the shapes. The moving shape sequences were then aligned to lie in the same reference space in both spatial and temporal dimensions. We employed an affine (rigid and isotropic scaling) spatial alignment. Another option would have been to use non-rigid spatial alignments. It would be interesting to investigate the effect of using different reference space definitions on the predicted motion outcome. The difference between the two approaches can be best seen when applying the same motion to two different shapes. In the affine (rigid) case the corresponding landmark points of the different shapes move in the same direction with the same magnitude. In the non-rigid case the motion vectors are deformed with the shape, and the resulting motion depends on the actual deformation applied to bring one shape into the other.

The created coronary artery motion models were employed for the intra-operative XA - 3D CTA registration in Chapter 7. In our experiments the fitted motion models performed at least as good as the 4D CTA based motion, and could therefore potentially substitute the 4D CTA derived motion. Before such a method can be applied in practice, there are several extensions to be considered. For example, one would desire a continuous overlay of the coronary arteries on the XA sequence for several heart beats. To achieve this, the sequence could be cut into separate heart beats in which the proposed method could be applied. A more sophisticated approach would be to use a sliding window to update the model parameters frame by frame. With enough latency built in the motion fitting, a good indicator for the respiratory phase, and a regular heart beat, the over-projection of the coronary arteries may even be possible in parts of the sequence with fading contrast. This was affirmed by the feasibility experiment in Chapter 7, where we used the fitted motion on a second heart beat with good results. With fading contrast, however, the reliability of the match would decrease. This reliability should then be provided together with the overlay to the surgeon.

8.2.2 Potential application areas

In this thesis we have proposed novel methods for SSM based dense shape reconstruction from sparse projection and point data. Though our experiments focused on specific anatomies and medical applications, the proposed methods are more generally applicable. In this section we briefly outline different application possibilities that may point to possible future research areas.

The first chapters focused on reconstruction and tracking of the femur. Despite this single focus, the same methodology is expected to work with other bones, such as the tibia, bones of the arm, fingers, ribs, etc., or soft tissues if sufficient contrast is available. 3D shape reconstruction from biplane X-rays is not only useful for preoperative planning, but could also provide surrogate measurements of e.g. cartilage thickness, by reconstructing the bones of a joint. 3D reconstruction could in that case minimize the acquisition angle dependency of the joint space measured on a 2D radiograph. Also, as a byproduct of the reconstruction, the proposed method performs segmentation of the 2D images. Such 3D model based segmentation may be interesting e.g. when a second projection can guide the segmentation of the X-ray of interest, or when the projected shape variation is non-linear (e.g. due to rotation) while the 3D shape variation is linear. Tracking of 3D shapes from fluoroscopy can be employed to measure bone and joint motion for instance for evaluation of surgical outcome of joint replacement surgeries. The post-operative motion is then compared either with the pre-operative motion, or with normal joint motion. Kinematic analyses of bones may also provide a surrogate measurement for e.g. ligament condition. Injured ligaments may result in abnormal bone motion patterns that can be detected with kinematic analysis [25, 128], even for ligaments so small that current imaging technologies do not allow their direct visualization. Also, relative bone motion could be used for inverse dynamic analyses estimating joint stress and muscle torques [41, 47]. Lastly, tracking of 3D shapes could also be useful for real-time visualization for image guided surgeries and interventions.

Current medical practice moves toward minimally invasive surgeries and interventions. Image guidance during these interventions is of major importance, as both the anatomy and the instruments are otherwise invisible to the surgeon. In Chapter 7 we focus on improved image guidance during percutaneous coronary interventions (PCI), by registering the coronary arteries extracted from a pre-operative 3D CTA image to the contrast enhanced intra-operative X-ray angiography (XA) sequences. This approach is useful for at least two reasons. First, in CTA the entire vascular tree is visible including segments after the occlusion and calcifications, as opposed to the XA sequences where the contrast agent cannot pass the occlusion. Second, as the intra-operative contrast agent is toxic, only limited number of sequences can be made during an intervention. Therefore, it would be useful to keep the registration to the CTA image also in low or no contrast sequences. This offers an interesting future research direction, in which a coronary motion model and a respiratory motion model could be combined with cardiac phase measurements, e.g. derived from the ECG signal, and respiratory measurements, e.g. by monitoring the motion of skin markers. In this case an estimation of the reliability of the over-projection is of major interest,

as the lack of contrast makes human assessment of the reliability not possible.

An advantage of model based reconstruction is the possibility to estimate confidence intervals (remaining uncertainties) on the estimated shapes, as shown in Chapter 5. Such indication of uncertain shape areas could be directly used for facilitating manual segmentation by guiding the user to shape areas most uncertain given the already annotated information. Uncertainties could also be used for theoretical selection of optimal scanning angles, or minimum amount of sparse information needed for a reconstruction of a shape with high confidence.

Bibliography

[1] E. S. Abebe, G. M. Utturkar, D. C. Taylor, C. E. Spritzer, J. P. Kim, C. T. M. III, W. E. Garrett, and L. E. DeFrate. The effects of femoral graft placement on in vivo knee kinematics after anterior cruciate ligament reconstruction. *Journal of Biomechanics*, 44(5):924–929, 2011.

[2] N. A. Ablitt, J. Gao, J. Keegan, L. Stegger, D. N. Firmin, and G.-Z. Yang. Predictive cardiac motion modeling and correction with partial least squares regression. *IEEE Trans. Med. Imaging*, 23(10):1315–1324, 2004.

[3] T. Albrecht, R. Knothe, and T. Vetter. Modeling the Remaining Flexibility of Partially Fixed Statistical Shape Models. In *Workshop on the Mathematical Foundations of Computational Anatomy*, 2008.

[4] W. Anderst, R. Zauel, J. Bishop, E. Demps, and S. Tashman. Validation of three-dimensional model-based tibio-femoral tracking during running. *Medical engineering & physics*, 31(1):10–6, Jan. 2009.

[5] N. Baka, M. de Bruijne, W. J. Niessen, J. H. C. Reiber, and B. P. F. Lelieveldt. Confidence of model based shape reconstruction from sparse data. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1077–1080, 2010.

[6] N. Baka, M. de Bruijne, T. van Walsum, B. Kaptein, J. Giphart, M. Schaap, W. Niessen, and B. Lelieveldt. Statistical Shape Model Based Femur Kinematics from Biplane Fluoroscopy. *IEEE Transactions on Medical Imaging*, accepted, Apr. 2012.

[7] N. Baka, B. L. Kaptein, M. de Bruijne, T. van Walsum, J. E. Giphart, W. J. Niessen, and B. P. F. Lelieveldt. 2D-3D shape reconstruction of the distal femur from stereo X-ray imaging using statistical shape models. *Medical Image Analysis*, 15(6):840–850, 2011.

[8] N. Baka, C. Metz, M. Schaap, B. P. F. Lelieveldt, W. J. Niessen, and M. de Bruijne. Comparison of Shape Regression Methods under Landmark Position Uncertainty. In G. Fichtinger, A. Martel, and T. Peters, editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2011*, volume 6892 of *Lecture Notes in Computer Science*, pages 434–441, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[9] N. Baka, W. Niessen, T. van Walsum, L. Ferrarini, J. H. C. Reiber, and
     B. Lelieveldt. Correspondence free 3D statistical shape model fitting to sparse
     x-ray projections. In *SPIE Medical Imaging*, pages 76230D1 – 76230D9, 2010.

[10] S. A. Banks and W. A. Hodge. Accurate measurement of three-dimensional
     knee replacement kinematics using single-plane fluoroscopy. *IEEE transactions
     on bio-medical engineering*, 43(6):638–49, June 1996.

[11] D. C. Barratt, C. S. K. Chan, P. J. Edwards, G. P. Penney, M. Slomczykowski,
     T. J. Carter, and D. J. Hawkes. Instantiation and registration of statistical
     shape models of the femur and pelvis using 3D ultrasound imaging. *Medical
     Image Analysis*, 12(3):358–374, 2008.

[12] S. Benameur, M. Mignotte, S. Parent, H. Labelle, W. Skalli, and J. D. Guise.
     3D/2D Registration and segmentation of scoliotic vertebrae using statistical
     models. *Computerized Medical Imaging and Graphics*, 27(5):321–337, 2003.

[13] J. Blackall, A. King, G. Penney, A. Adam, D. Hawkes, W. Niessen, and
     M. Viergever. A Statistical Model of Respiratory Motion and Deformation
     of the Liver. In *Medical Image Computing and Computer-Assisted Interven-
     tion MICCAI 2001*, volume 2208 of *Lecture Notes in Computer Science*, pages
     1338–1340, Berlin, Heidelberg, Oct. 2001. Springer Berlin Heidelberg.

[14] R. Blanc, M. Reyes, C. Seiler, and G. Szǧż″kely. Conditional variability of sta-
     tistical shape models based on surrogate variables. In *Medical Image Computing
     and Computer-Assisted Intervention MICCAI 2009*, 2009.

[15] R. Blanc, C. Seiler, G. Székely, L.-P. Nolte, and M. Reyes. Statistical Model
     Based Shape Prediction from a Combination of Direct Observations and Various
     Surrogates. Application to Orthopaedic Research. *Medical Image Analysis*, May
     2012.

[16] R. Blanc, E. Syrkina, and G. Székely. Estimating the confidence of statistical
     model based shape prediction. In *IPMI*, 5636, pages 602–613, 2009.

[17] R. Blanc and G. Szekely. Confidence regions for statistical model based shape
     prediction from sparse observations. *IEEE transactions on medical imaging*,
     Feb. 2012.

[18] B. P. Boden, G. S. Dean, J. A. Feagin, and W. E. Garrett. Mechanisms of
     anterior cruciate ligament injury. *Orthopedics*, 23(6):573–8, June 2000.

[19] J. G. Bosch, S. C. Mitchell, B. P. F. Lelieveldt, F. Nijland, O. Kamp, M. Sonka,
     and J. H. C. Reiber. Automatic segmentation of echocardiographic sequences
     by active appearance motion models. *IEEE Transactions on Medical Imaging*,
     21(11):1374–83, Nov. 2002.

[20] S. Bouattour, R. Arndt, and D. Paulus. 4D Reconstruction of Coronary Arteries from Monoplane Angiograms. In A. Gagalowicz and W. Philips, editors, *Computer Analysis of Images and Patterns*, volume 3691 of *Lecture Notes in Computer Science*, pages 724–731–731, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[21] J. Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, Nov. 1986.

[22] J. Cates, P. T. Fletcher, M. Styner, M. Shenton, and R. Whitaker. Shape modeling and analysis with entropy-based particle systems. *Information processing in medical imaging : proceedings of the … conference*, 20:333–45, Jan. 2007.

[23] R. Chandrashekara, A. Rao, G. Sanchez-Ortiz, R. Mohiaddin, and D. Rueckert. Construction of a Statistical Model for Cardiac Motion Analysis Using Nonrigid Image Registration. In C. Taylor and J. A. Noble, editors, *Information Processing in Medical Imaging*, volume 2732 of *Lecture Notes in Computer Science*, pages 599–610, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[24] X. Chen, J. Graham, C. Hutchinson, and L. Muir. Inferring 3D Kinematics of Carpal Bones from Single View Fluoroscopic Sequences. In G. Fichtinger, A. L. Martel, and T. M. Peters, editors, *MICCAI (2)*, volume 6892 of *Lecture Notes in Computer Science*, pages 680–687. Springer, 2011.

[25] X. Chen, J. Graham, C. Hutchinson, and L. Muir. Inferring 3D Kinematics of Carpal Bones from Single View Fluoroscopic Sequences. In G. Fichtinger, A. Martel, and T. Peters, editors, *MICCAI 2011, Medical Image Computing and Computer-Assisted Intervention*, volume 6892 of *Lecture Notes in Computer Science*, pages 680–687, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[26] D. Comaniciu, P. Meer, and S. Member. Mean shift: A robust approach toward feature space analysis. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 24:603 – 619, 2002.

[27] T. F. Cootes, D. H. Cooper, C. J. Taylor, and J. Graham. Trainable method of parametric shape description. *Image Vision Comput.*, 10(5):289–294, 1992.

[28] T. F. Cootes, G. Edwards, and C. J. Taylor. Comparing active shape models with active appearance models. In *BMVC*, Aug. 1999.

[29] T. F. Cootes and C. J. Taylor. Data Driven Refinement of Active Shape Model Search. In *Proc. British Machine Vision Conf.*, pages 383–392, 1996.

[30] T. F. Cootes and C. J. Taylor. Constrained Active Appearance Models. *Computer Vision, IEEE International Conference on*, 1:748, 2001.

[31] T. F. Cootes and C. J. Taylor. Statistical Models of Appearance for Computer Vision. Technical report, University of Manchester, t.cootes@man.ac.uk, Mar. 2004.

[32] T. Cresson, D. Branchaud, R. Chav, B. Godbout, and J. A. de Guise. 3D shape reconstruction of bone from two x-ray images using 2D/3D non-rigid registration based on moving least-squares deformation. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7623, pages 76230F–76230F–9, Mar. 2010.

[33] 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. *IEEE transactions on medical imaging*, 21(5):525–37, May 2002.

[34] M. de Bruijne, M. T. Lund, L. B. Tankó, P. C. Pettersen, and M. Nielsen. Quantitative vertebral morphometry using neighbor-conditional shape models. *Med. Image Anal.*, 11(5):503–512, 2007.

[35] M. de Bruijne, B. van Ginneken, M. A. Viergever, and W. J. Niessen. Adapting Active Shape Models for 3D segmentation of tubular structures in medical images. *Information processing in medical imaging : proceedings of the ... conference*, 18:136–47, July 2003.

[36] D. A. Dennis, M. R. Mahfouz, R. D. Komistek, and W. Hoff. In vivo determination of normal and anterior cruciate ligament-deficient knee kinematics. *Journal of biomechanics*, 38(2):241–53, Feb. 2005.

[37] C. Di Mario, G. S. Werner, G. Sianos, A. R. Galassi, J. Büttner, D. Dudek, B. Chevalier, T. Lefevre, J. Schofer, J. Koolen, H. Sievert, B. Reimers, J. Fajadet, A. Colombo, A. Gershlick, P. W. Serruys, and N. Reifart. European perspective in the recanalisation of Chronic Total Occlusions (CTO): consensus document from the EuroCTO Club. *EuroIntervention : journal of EuroPCR in collaboration with the Working Group on Interventional Cardiology of the European Society of Cardiology*, 3(1):30–43, May 2007.

[38] J. Dworzak, H. Lamecker, J. von Berg, T. Klinder, C. Lorenz, D. Kainmüller, H. Seim, H.-C. Hege, and S. Zachow. 3D reconstruction of the human rib cage from 2D projection images using a statistical shape model. *International journal of computer assisted radiology and surgery*, 5(2):111–24, Mar. 2010.

[39] J. Ehrhardt, R. Werner, A. Schmidt-Richberg, and H. Handels. Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration. *IEEE Transactions on Medical Imaging*, 30(2):251–265, 2011.

[40] J. Feldmar, N. Ayache, and F. Betting. 3D-2D Projective Registration of Free-Form Curves and Surfaces. In *International Conference on Computer Vision ICCV' 95*, pages 549–556, 1995.

[41] J. W. Fernandez, M. Akbarshahi, H. J. Kim, and M. G. Pandy. Integrating modelling, motion capture and x-ray fluoroscopy to investigate patellofemoral function during dynamic activity. *Computer methods in biomechanics and biomedical engineering*, 11(1):41–53, Feb. 2008.

[42] L. Ferrarini, H. Olofsen, W. M. Palm, M. A. van Buchem, J. H. Reiber, and F. Admiraal-Behloul. GAMEs: growing and adaptive meshes for fully automatic shape modeling and analysis. *Medical Image Analysis*, 11:302–314, 2007.

[43] M. Fleute and S. Lavallée. Building a Complete Surface Model from Sparse Data Using Statistical Shape Models: Application to Computer Assisted Knee Surgery System. In *MICCAI '98: Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 879–887, London, UK, 1998. Springer-Verlag.

[44] M. Fleute and S. Lavallée. Nonrigid 3-D/2-D Registration of Images Using Statistical Models. In *MICCAI '99: Proceedings of the Second International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 138–147, London, UK, 1999. Springer-Verlag.

[45] M. Fleute, S. Lavallée, and L. Desbat. Integrated Approach for Matching Statistical Shape Models with Intra-operative 2D and 3D Data. In *MICCAI '02: Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention-Part II*, pages 364–372, London, UK, 2002. Springer-Verlag.

[46] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever. Multiscale vessel enhancement filtering. In *Medical Image Computing and Computer-Assisted Interventation MICCAI 98*, volume 1496, pages 130 – 137, 1998.

[47] B. J. Fregly, H. A. Rahman, and S. A. Banks. Theoretical Accuracy of Model-Based Shape Matching for Measuring Natural Knee Kinematics with Single-Plane Fluoroscopy. *Journal of Biomechanical Engineering*, 127(4):692–699, Aug. 2005.

[48] P. Gamage, S. Q. Xie, P. Delmas, and P. Xu. 3D Reconstruction of Patient Specific Bone Models from 2D Radiographs for Image Guided Orthopedic Surgery. In *Proceedings of the 2009 Digital Image Computing: Techniques and Applications*, DICTA '09, pages 212–216, Washington, DC, USA, 2009. IEEE Computer Society.

[49] W. F. Ganong. *Review of Medical Physiology*. Lange Medical Books/McGraw-Hill, 22nd editi edition, 2005.

[50] C. Gatta, S. Balocco, V. Martin-Yuste, R. Leta, and P. Radeva. Non-rigid Multi-modal Registration of Coronary Arteries Using SIFTflow. In J. Vitrià, J. a. M. Sanches, and M. Hernández, editors, *Pattern Recognition and Image Analysis*, volume 6669 of *Lecture Notes in Computer Science*, pages 159–166–166, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[51] M. Groher, D. Zikic, and N. Navab. Deformable 2D-3D registration of vascular structures in a one view scenario. *IEEE Transactions on Medical Imaging*, 28(6):847–60, June 2009.

[52] O. L. A. Harrysson, Y. A. Hosni, and J. F. Nayfeh. Custom-designed orthopedic implants evaluated using finite element analysis of patient-specific computed tomography data: femoral-component case study. *BMC Musculoskeletal Disorders*, 8:91, 2007.

[53] R. Hart, P. ďž″vďž″b, and P. Filan. Intraoperative navigation in hip surface arthroplasty: a radiographic comparative analysis study. *Archives of Orthopaedic and Trauma Surgery*, 128(4):429–434, Apr. 2008.

[54] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning (Data mining, Inference, and Prediction)*. Springer, 2009.

[55] T. Heimann and H.-P. Meinzer. Statistical shape models for 3D medical image segmentation: A review. *Medical Image Analysis*, 2009.

[56] G. Heitz, T. Rohlfing, and C. R. Maurer Jr. Statistical shape model generation using nonrigid deformation of a template mesh. In J. M. Fitzpatrick and J. M. Reinhardt, editors, *Medical Imaging 2005: Image Processing*, volume 5747 of *Proceedings of SPIE*, pages 1411–1421, Bellingham, WA, 2005. The International Society for Optical Engineering, SPIE Press.

[57] A. Hill, T. F. Cootes, and C. J. Taylor. Active Shape Models and the Shape Approximation Problem. In *British Machine Vision Conference*, 1995.

[58] C. Hoogendoorn, F. Sukno, S. Ordas, and A. F. Frangi. Bilinear Models for Spatio-Temporal Point Distribution Analysis. *International Journal of Computer Vision*, 85(3):237–252, 2009.

[59] H. Hufnagel, X. Pennec, J. Ehrhardt, N. Ayache, and H. Handels. A Global Criterion for the Computation of Statistical Shape Model Parameters Based on Correspondence Probabilities. In *Bildverarbeitung fďž″r die Medizin 2008*. Springer Berlin Heidelberg, 2008.

[60] A. Hurvitz and L. Joskowicz. Registration of a CT-like atlas to fluoroscopic X-ray images using intensity correspondences. *International Journal of Computer Assisted Radiology and Surgery (CARS)*, 3:493–504, 2008.

[61] Y. Iwashita, R. Kurazume, T. Hasegawa, and K. Hara. Fast Alignment of 3D Geometrical Models and 2D Color Images Using 2D Distance Maps. In *3DIM'05: Fifth International Conference on 3-D Digital Imaging and Modeling*, pages 164–171, 2005.

[62] M. d. B. J. Iglesias. Semiautomatic Segmentation of Vertebrae in Lateral X-rays Using a Conditional Shape Model. *Academic Radiology*, 14(10):1156–1165, 2007.

[63] A. K. Jain, Y. Zhong, and S. Lakshmanan. Object Matching Using Deformable Templates. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(3):267–278, 1996.

[64] B. L. Kaptein, K. B. Shelburne, M. R. Torry, and J. E. Giphart. A comparison of calibration methods for stereo fluoroscopic imaging systems. *Journal of biomechanics*, 44(13):2511–5, Sept. 2011.

[65] B. L. Kaptein, E. R. Valstar, B. C. Stoel, P. M. Rozing, and J. H. C. Reiber. A new model-based RSA method validated using CAD models and models from reversed engineering. *J Biomech*, 36(6):873–882, 2003.

[66] H. A. Kirisli, M. Schaap, S. Klein, S. L. Papadopoulou, M. Bonardi, C. H. Chen, A. C. Weustink, N. R. A. Mollet, E. P. A. Vonken, R. J. van der Geest, T. van Walsum, and W. J. Niessen. Evaluation of a multi-atlas based method for segmentation of cardiac CTA data: a large-scale, multi-center and multi-vendor study. *Medical Physics*, 37(12):6279–6292, 2010.

[67] Y. Kita, D. Wilson, J. Noble, W. Wells, A. Colchester, and S. Delp. Real-time registration of 3D cerebral vessels to X-ray angiograms. In W. M. Wells, A. Colchester, and S. Delp, editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 98*, volume 1496 of *Lecture Notes in Computer Science*, pages 1125–1133, Berlin/Heidelberg, 1998. Springer-Verlag.

[68] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim. elastix: a toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205, Jan. 2010.

[69] T. Klinder, C. Lorenz, and J. Ostermann. Prediction framework for statistical respiratory motion modeling. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2010*, MICCAI'10, pages 327–334, Berlin, Heidelberg, 2010. Springer-Verlag.

[70] A. C. Kotcheff and C. J. Taylor. Automatic construction of eigenshape models by direct optimization. *Medical Image Analysis*, 2(4):303–314, Dec. 1998.

[71] P. R. Krekel, C. P. Botha, E. R. Valstar, P. W. de Bruin, P. M. Rozing, and F. H. Post. Interactive simulation and comparative visualisation of the bone-determined range of motion of the human shoulder. In *SimVis*, pages 275–288, 2006.

[72] R. Kurazume, K. Nakamura, T. Okada, Y. Sato, N. Sugano, T. Koyama, Y. Iwashita, and T. Hasegawa. 3D reconstruction of a femoral shape using a parametric model and two 2D fluoroscopic images. *Comput. Vis. Image Underst.*, 113(2):202–211, 2009.

[73] H. Lamecker, T. H. Wenckebach, and H.-C. Hege. Atlas-based 3D-Shape Reconstruction from X-Ray Images. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pages 371–374, Washington, DC, USA, 2006. IEEE Computer Society.

[74] S. Laporte, W. Skalli, J. A. de Guise, F. Lavaste, and D. Mitton. A Biplanar Reconstruction Method Based on 2D and 3D Contours: Application to the Distal Femur. *Computer Methods in Biomechanics and Biomedical Engineering*, 6:1–6, 2003.

[75] S. Lavallée, R. Szeliski, and L. Brunie. Matching 3-D Smooth Surfaces with their 2-D Projections using 3-D Distance Maps. In *Selected Papers from the Workshop on Geometric Reasoning for Perception and Action*, pages 217–238, London, UK, 1993. Springer-Verlag.

[76] J. Liu and J. K. Udupa. Oriented active shape models. *IEEE transactions on medical imaging*, 28(4):571–84, Apr. 2009.

[77] T. Liu, D. Shen, and C. Davatzikos. Predictive Modeling of Anatomic Structures Using Canonical Correlation Analysis. In *IEEE International Symposium on Biomedical Imaging: Nano to Macro - ISBI*, pages 1279–1282, Apr. 2004.

[78] X. Liu, I. Oguz, S. M. Pizer, and G. S. Mageras. Shape-correlated deformation statistics for respiratory motion prediction in 4D lung. In *Proceedings of the SPIE*, volume 7625, 2010.

[79] H. Livyatan, Z. Yaniv, S. Member, L. Joskowicz, and S. Member. Gradient-based 2-D/3-D rigid registration of fluoroscopic X-ray to CT. *IEEE Trans. Med. Imag*, 22:1395–1406, 2003.

[80] C. Lorenz and J. von Berg. A comprehensive shape model of the heart. *Medical Image Analysis*, 10(4):657–70, Aug. 2006.

[81] J. Lötjönen, K. Antila, E. Lamminmäki, J. Koikkalainen, M. Lilja, and T. F. Cootes. Artificial Enlargement of a Training Set for Statistical Shape Models: Application to Cardiac Images. In *FIMH*, pages 92–101, 2005.

[82] P. A. Lotke. *Master Techniques in Orthopaedic Surgery: Knee Arthroplasty*. Lippincott Williams & Wilkins, 2008.

[83] M. Magro, C. Schultz, C. Simsek, H. Garcia-Garcia, E. Regar, K. Nieman, N. Mollet, P. Serruys, and R. van Geuns. Computed tomography as a tool for percutaneous coronary intervention of chronic total occlusions. *EuroIntervention*, pages G123–31, 2010.

[84] P. Markelj, D. Tomaževič, B. Likar, and F. Pernuš. A review of 3{D}/2{D} registration methods for image-guided interventions. *Medical Image Analysis*, 16(3):642–61, 2010.

[85] A. Mazzoli, M. Germani, and R. Raffaeli. Direct fabrication through electron beam melting technology of custom cranial implants designed in a PHANToM-based haptic environment. *Materials & Design*, 30(8):3186–3192, 2009.

[86] R. A. McLaughlin, J. Hipwell, D. J. Hawkes, J. A. Noble, J. V. Byrne, and T. C. S. Cox. A Comparison of 2D-3D Intensity-Based Registration and Feature-Based Registration for Neurointerventions. In *MICCAI '02 Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 517–524, Sept. 2002.

[87] K. McLeish, D. L. G. Hill, D. Atkinson, J. M. Blackall, and R. Razavi. A study of the motion and deformation of the heart due to respiration. *IEEE Transactions on Medical Imaging*, 21(9):1142–50, Sept. 2002.

[88] C. Metz, N. Baka, H. A. Kirisli, M. Schaap, S. Klein, L. Neefjes, N. Mollet, B. Lelieveldt, M. de Bruijne, W. Niessen, and T. van Walsum. Regression-based cardiac motion prediction from single-phase CTA. *IEEE Transactions on Medical Imaging*, 31(6):1311–1325, 2012.

[89] C. Metz, M. Schaap, S. Klein, L. Neefjes, E. Capuano, C. Schultz, R. van Geuns, P. Serruys, T. van Walsum, W. Niessen, G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor. Patient specific 4D coronary models from ECG-gated CTA data for intra-operative dynamic alignment of CTA with X-ray images. In G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2009*, volume 5761 of *Lecture Notes in Computer Science*, pages 369–376, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[90] C. Metz, M. Schaap, S. Klein, P. Rijnbeek, L. Neefjes, N. Mollet, C. Schultz, P. Serruys, W. Niessen, and T. van Walsum. Alignment of 4D Coronary CTA with Monoplane X-Ray Angiography. In *MICCAI workshop: Augmented Environments for Computer assisted interventions*, 2011.

[91] C. Metz, M. Schaap, S. Klein, A. Weustink, N. Mollet, C. Schultz, R. van Geuns, P. Serruys, T. van Walsum, and W. Niessen. GPU accelerated alignment of 3-D CTA with 2-D X-ray data for improved guidance in coronary interventions. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 959–962. IEEE, June 2009.

[92] C. T. Metz, N. Baka, H. A. Kirisli, M. Schaap, T. van Walsum, S. Klein, L. Neefjes, N. R. A. Mollet, B. P. F. Lelieveldt, M. de Bruijne, and W. J. Niessen. Conditional shape models for cardiac motion estimation. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2010, Part I, LNCS 6361*, pages 452–459, 2010.

[93] C. T. Metz, M. Schaap, A. C. Weustink, N. R. Mollet, T. van Walsum, and W. J. Niessen. Coronary centerline extraction from CT coronary angiography images using a minimum cost path approach. *Medical Physics*, 36(12):5568, Nov. 2009.

[94] D. L. Miranda, M. J. Rainbow, E. L. Leventhal, J. J. Crisco, and B. C. Flem-
      ing.   Automatic determination of anatomical coordinate systems for three-
      dimensional bone models of the isolated human knee. *Journal of Biomechanics*,
      43(8):1623–1626, 2010.

[95] S. C. Mitchell, B. P. F. Lelieveldt, R. J. van der Geest, H. G. Bosch, J. H. C.
      Reiber, and M. Sonka. Multistage Hybrid Active Appearance Model Match-
      ing: Segmentation of Left and Right Ventricles in Cardiac MR Images. *IEEE
      Transactions on Medical Imaging*, 20(5):415–423, 2001.

[96] K. L. Moore, A. F. Dalley, and A. M. R. Agur. *Clinically Oriented Anatomy*.
      Lippincott Williams & Wilkins, 6th editio edition, 2008.

[97] F. Mourgues. 3D+t Modeling of Coronary Artery Tree from Standard Non
      Simultaneous Angiograms. *Pattern Recognition*, 2208:1320–1322, 2001.

[98] E. Oost, G. Koning, M. Sonka, P. V. Oemrawsingh, J. H. C. Reiber, and B. P. F.
      Lelieveldt. Automatec contour detection in X-ray left ventricular angiograms
      using multiview active appearance models and dynamic programming. *IEEE
      Trans. Med. Imaging*, 25:1158–1171, 2006.

[99] S. Ordas, E. Oubel, R. Leta, F. Carreras, and A. F. Frangi. A statistical
      shape model of the heart and its application to model-based segmentation. In
      Proceedings of the SPIE, volume 6511 of *Presented at the Society of Photo-
      Optical Instrumentation Engineers (SPIE) Conference*, Mar. 2007.

[100] R. R. Paulsen and K. B. Hilger. Shape modelling using Markov random field
      restoration of point correspondences. *Information processing in medical imaging
      : proceedings of the … conference*, 18:1–12, July 2003.

[101] G. P. Penney, P. G. Batchelor, D. L. G. Hill, D. J. Hawkes, and J. Weese. Val-
      idation of a two- to three-dimensional registration algorithm for aligning pre-
      operative CT images and intraoperative fluoroscopy images. *Medical Physics*,
      28(6):1024–1032, 2001.

[102] D. Perperidis, R. Mohiaddin, D. Rueckert, J. Duncan, and G. Gerig. Construc-
      tion of a 4D Statistical Atlas of the Cardiac Anatomy and Its Use in Classifi-
      cation. In J. S. Duncan and G. Gerig, editors, *Medical Image Computing and
      Computer-Assisted Intervention MICCAI 2005*, volume 3750 of *Lecture Notes
      in Computer Science*, pages 402–410, Berlin, Heidelberg, 2005. Springer Berlin
      Heidelberg.

[103] S. M. Pizer, D. S. Fritsch, P. A. Yushkevich, V. E. Johnson, and E. L. Chaney.
      Segmentation, registration, and measurement of shape variation via image ob-
      ject shape. *IEEE transactions on medical imaging*, 18(10):851–65, Oct. 1999.

[104] K. T. Rajamani, M. A. Styner, H. Talib, G. Zheng, L. P. Nolte, and M. A.
      Ballester. Statistical deformable bone models for robust 3D surface extrapola-
      tion from sparse data. *Medical Image Analysis*, 11(2):99–109, Apr. 2007.

[105] A. Rao, P. Aljabar, and D. Rueckert. Hierarchical statistical shape analysis and prediction of sub-cortical brain structures. *Medical Image Analysis*, 12(1):55–68, 2008.

[106] A. Rao, K. O. Babalola, and D. Rueckert. Canonical Correlation Analysis of Sub-cortical Brain Structures Using Non-rigid Registration. In *WBIR*, pages 66–74, 2006.

[107] M. S. Reis and P. M. Saraiva. A comparative study of linear regression methods in noisy environments. *J. Chemometrics*, 18:526–536, 2004.

[108] D. Ruijters, B. M. ter Haar Romeny, and P. Suetens. Vesselness-based 2D-3D registration of the coronary arteries. *International journal of computer assisted radiology and surgery*, 4(4):391–7, June 2009.

[109] O. Sadowsky, G. Chintalapani, and R. H. Taylor. Deformable 2D-3D Registration of the Pelvis with a Limited Field of View, Using Shape Statistics. In *MICCAI (2)*, pages 519–526, 2007.

[110] M. Schaap, T. van Walsum, L. Neefjes, C. Metz, E. Capuano, M. de Bruijne, and W. Niessen. Robust shape regression for supervised vessel segmentation and its application to coronary segmentation in CTA. *IEEE Transactions on Medical Imaging*, 30(11):1974–86, Nov. 2011.

[111] E. Serradell, A. Romero, R. Leta, C. Gatta, and F. Moreno-Noguer. Simultaneous correspondence and non-rigid 3D reconstruction of the coronary tree from single X-ray images. In *International Conference on Computer Vision ICCV2011*, pages 850–857, 2011.

[112] G. Shechter, C. Ozturk, J. R. Resar, and E. R. McVeigh. Respiratory motion of the heart from free breathing coronary angiograms. *IEEE transactions on medical imaging*, 23(8):1046–56, Aug. 2004.

[113] G. Shechter, J. R. Resar, and E. R. McVeigh. Displacement and velocity of the coronary arteries: cardiac and respiratory motion. *IEEE Transactions on Medical Imaging*, 25(3):369–75, Mar. 2006.

[114] R. A. Siston, N. J. Giori, S. B. Goodman, and S. L. Delp. Surgical navigation for total knee arthroplasty: A perspective. *Journal of Biomechanics*, 40(4):728–735, 2007.

[115] K. H. Soon, J. B. Selvanayagam, N. Cox, A.-M. Kelly, K. W. Bell, and Y. L. Lim. Percutaneous revascularization of chronic total occlusions: review of the role of invasive and non-invasive imaging modalities. *International Journal of Cardiology*, 116(1):1–6, Mar. 2007.

[116] M. Styner, G. Gerig, J. Lieberman, D. Jones, and D. Weinberger. Statistical shape analysis of neuroanatomical structures based on medial models. *Medical Image Analysis*, 7(3):207–220, Sept. 2003.

[117] A. Suinesiaputra, A. F. Frangi, T. A. M. Kaandorp, H. J. Lamb, J. J. Bax, J. H. C. Reiber, and B. P. F. Lelieveldt. Automated detection of regional wall motion abnormalities based on a statistical model applied to multislice short-axis cardiac MR images. *IEEE transactions on medical imaging*, 28(4):595–607, May 2009.

[118] G. Székely, A. Kelemen, C. Brechbühler, and G. Gerig. Segmentation of 2-D and 3-D objects from MRI volume data using constrained elastic deformations of flexible Fourier contour and surface models. *Medical image analysis*, 1(1):19–34, Mar. 1996.

[119] T. S. Y. Tang and R. E. Ellis. 2D/3D deformable registration using a hybrid atlas. *Med Image Comput Comput Assist Interv*, 8(Pt 2):223–230, 2005.

[120] M. Taron, N. Paragios, and M.-P. Jolly. Uncertainty-Driven Non-parametric Knowledge-Based Segmentation: The Corpus Callosum Case. In *VLSM*, pages 198–209, 2005.

[121] D. Tomazevic, B. Likar, and F. Pernus. Gradient-based registration of 3D MR and 2D X-ray images. In H. U. Lemke, M. W. Vannier, K. Inamura, A. G. Farman, and K. Doi, editors, *CARS*, volume 1230 of *International Congress Series*, pages 338–345. Elsevier, 2001.

[122] M. R. Torry, K. B. Shelburne, D. S. Peterson, J. E. Giphart, J. P. Krong, C. Myers, J. R. Steadman, and S. L.-Y. Woo. Knee Kinematic Profiles During Drop Landings: A Biplane Fluoroscopy Study. *Med Sci Sports Exerc*, 43(3):533–41, 2010.

[123] R. Toth and A. Madabhushi. Multi-Feature Landmark-Free Active Appearance Models: Application to Prostate MRI Segmentation. *IEEE transactions on medical imaging*, May 2012.

[124] G.-A. Turgeon, G. Lehmann, G. Guiraudon, M. Drangova, D. Holdsworth, and T. Peters. 2D-3D registration of coronary angiograms for cardiac procedure planning and guidance. *Medical Physics*, 32(12):3737, 2005.

[125] C. J. Twining, C. J. Twining, and C. J. Taylor. Kernel Principal Component Analysis and the Construction Of Non-Linear Active Shape Models. *IN PROC. BMVC*, pages 23 – 32, 2001.

[126] E. R. Valstar, R. Gill, L. Ryd, G. FLivik, N. Börlin, and J. Kärrholm. Guidelines for standadrization of radiostereogrammetry (RSA) of implants. *Acta Orthop.*, 76(4):563–572, 2005.

[127] H. C. van Assen, M. G. Danilouchkine, A. F. Frangi, S. Ordas, J. J. Westenberg, J. H. Reiber, and B. P. Lelieveldt. SPASM: A 3D-ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data. *Medical Image Analysis*, 10(2):286–303, Apr. 2006.

[128] M. van de Giessen, M. Foumani, F. M. Vos, S. D. Strackee, M. Maas, L. J. Van Vliet, C. A. Grimbergen, and G. J. Streekstra. A 4D statistical model of wrist bone motion patterns. *IEEE transactions on medical imaging*, 31(3):613–25, Mar. 2012.

[129] E. B. van de Kraats, G. P. Penney, D. Tomazevic, T. van Walsum, and W. J. Niessen. Standardized Evaluation Methodology for 2D-3D Registration. *IEEE Transactions on Medical Imaging*, 24(9):1177–1190, 2005.

[130] B. van Ginneken, M. B. Stegmann, and M. Loog. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. *Medical image analysis*, 10(1):19–40, Feb. 2006.

[131] P. D. Wentzell, D. T. Andrews, D. C. Hamilton, F. Klaas, and B. R. Kowalski. Maximum likelihood principal component analysis. *Journal of Chemometrics*, 11:339–366, July 1997.

[132] T. Whitmarsh, L. Humbert, M. De Craene, L. M. Del Rio Barquero, and A. F. Frangi. Reconstructing the 3D shape and bone mineral density distribution of the proximal femur from dual-energy X-ray absorptiometry. *IEEE transactions on medical imaging*, 30(12):2101–14, Dec. 2011.

[133] Y. M. Yang, D. Rueckert, and A. M. J. Bull. Predicting the Shapes of Bones at a Joint: Application to the Shoulder. *Computer Methods in Biomechanics and Biomedical Engineering*, 11(1):19–30, Feb. 2008.

[134] J. Yao and R. Taylor. Assessing Accuracy Factors in Deformable 2D/3D Medical Image Registration Using a Statistical Pelvis Model. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, page 1329, Washington, DC, USA, 2003. IEEE Computer Society.

[135] B. M. You, P. Siy, W. Anderst, and S. Tashman. In vivo measurement of 3-D skeletal kinematics from sequences of biplane radiographs: application to knee kinematics. *IEEE transactions on medical imaging*, 20(6):514–25, June 2001.

[136] H. Zhang, A. Wahle, R. K. Johnson, T. D. Scholz, and M. Sonka. 4-D cardiac MR image analysis: left and right ventricular morphology and function. *IEEE Transactions on Medical Imaging*, 29(2):350–64, Feb. 2010.

[137] G. Zheng. Personalized X-ray reconstruction of the proximal femur via intensity-based non-rigid 2D-3D registration. In *Medical image computing and computer-assisted intervention : MICCAI 2011, International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 14, pages 598–606, Jan. 2011.

[138] G. Zheng, X. Dong, K. T. Rajamani, X. Zhang, M. Styner, R. U. Tho-ranaghatte, L.-P. Nolte, and M. A. G. Ballester. Accurate and Robust Reconstruction of a Surface Model of the Proximal Femur from Sparse Point Data

and a Dense Point Distribution Model for Surgical Navigation. In *Biomedical Engineering, IEEE Transactions on*, volume 54 of *12*, pages 2109–2122, Dec. 2007.

[139] G. Zheng, S. Gollmer, S. Schumann, X. Dong, T. Feilkas, and M. A. G. Ballester. A 2D/3D correspondence building method for reconstruction of a patient-specific 3D bone surface model using point distribution models and calibrated X-ray images. *Medical Image Analysis*, 13(6):883–899, 2008.

[140] G. Zheng and L.-P. Nolte. Surface Reconstruction of Bone from X-ray Images and Point Distribution Model Incorporating a Novel Method for 2D-3D Correspondence. In *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2237–2244, Washington, DC, USA, 2006. IEEE Computer Society.

[141] G. Zheng and S. Schumann. 3D reconstruction of a patient-specific surface model of the proximal femur from calibrated x-ray radiographs: A validation study. *Medical Physics*, 36(4):1155–1166, 2009.

[142] Z. Zhu and G. Li. Construction of 3D human distal femoral surface models using a 3D statistical deformable model. *Journal of biomechanics*, 44(13):2368–2362, July 2011.

Layman's summary

In current clinical practice, medical imaging plays a key role in diagnosis, therapy planning and therapy monitoring. In most diagnostic settings imaging provides high resolution volumetric anatomical and/or functional information. However, in certain practical situations, limitations with respect to imaging time, space, radiation dose, or ergonomics make it impossible to acquire such rich data. In such cases, imaging may be performed that is of lower dimensionality than the desired information, or is sparse in at least one of the dimensions.

In this thesis we proposed to complement sparse image data with prior knowledge about the desired information to reconstruct it in its native dimensionality. The desired information was either a shape, or a shape and its motion over time. Prior shape knowledge was learned from a set of training shapes, or training shapes and their motion over time. Such a training set allows the construction of a model consisting of a population mean and main modes of variation. The resulting models are termed statistical shape models (SSM) and statistical motion models (SMM) respectively.

We applied shape reconstruction from sparse data in several medical applications. In Chapter 2 we proposed a method for reconstructing the femur (thigh bone) from two calibrated X-ray images. This is achieved by minimizing the distance between a SSM of the femur, and edges in the X-ray images. The 3D bone surface is required for pre-operative surgery planning and intra-operative guidance in recent computer guided surgical systems. The bone surfaces can be extracted from CT images, however, X-ray based reconstruction would be beneficial for both patient and hospital, as it decreased radiation dose and costs. We tested our method in ten in-vivo datasets, on which the achieved accuracy was 1.68 mm. This accuracy is comparable to other state of the art reconstruction methods, with the advantage that our method requires less manual interaction.

In Chapter 3 we further extended this method with the aim to reconstruct and track the femur bone from two calibrated X-ray fluoroscopic sequences (X-ray movies). Such fluoroscopic sequences are employed for accurate motion estimation of joints, e.g. for studying ligament function or assessing operative outcome. Current motion analysis methods align the patient specific bone shapes extracted from prior 3D image acquisitions, with the fluoroscopic sequence frame by frame. We propose to eliminate the 3D anatomy acquisition, lowering costs and potentially radiation dose. Furthermore, we enable automated tracking with a single manual initialization, which is expected to drastically reduce operator time. We show that shape reconstruction from

the entire sequence outperforms single frame reconstruction. The method achieved an average point-to-surface shape accuracy of 1.48 mm.

The second part of the work focused on the cardiac anatomy and motion. In Chapter 4 we proposed a method to predict patient specific heart motion from its shape derived from a 3D CT angiography (CTA) image. Hereto we performed linear regression between heart shape and its motion, based on a training set of 3D+time heart segmentations obtained from dynamic CT. The predicted motion is expected to help overlaying the coronary artery tree extracted from a 3D CTA image on X-ray angiographic sequences for improved guidance during minimally invasive percutaneous coronary intervention (PCI).

The cardiac segmentations from CTA images were achieved with a certain level of uncertainty. Image segmentations and point detections are indeed never exact, due to among others noise and finite resolution of medical images. Therefore, in Chapter 5 we derived a maximum aposteriori rigde regression scheme for shape prediction, taking the uncertainty of the segmentation into account. A thorough evaluation in Chapter 6 showed, that the proposed method outperformed other regression methods when the uncertainty in the test data is considerably different from the training set.

Finally, in Chapter 7 we combined the 2D/3D reconstruction work of Chapter 2 and 3 with the 3D-4D prediction work of Chapter 4, and investigated the applicability of statistical coronary artery motion models for the alignment of the coronary arteries with intra-operative X-ray angiography sequences. This alignment is desired for improved guidance during PCI procedures, as the arterial tree extracted from pre-operative 3D CTA images contain information not visible in the X-ray images of patients with a total occlusion. Such additional information include the vessel parts after the occlusion site and calcifications. We found that the 2D/3D alignment through time using statistical motion models perform at least as good as the patient specific coronary motion extracted from a 4D CTA image. These results are encouraging, as the number of truly 3D CTA acquisitions is increasing due to radiation dose considerations.

Samenvatting

In de klinische praktijk vervult medische beeldvorming vandaag de dag een sleutelrol bij het stellen van een diagnose, het plannen van een behandeling en het beoordelen van de effectiviteit van een behandeling. In de meeste gevallen zorgt beeldvorming voor volumetrische anatomische en/of functionele informatie met een hoge resolutie. In de praktijk maken beperkingen met betrekking tot acquisitietijd, ruimte, stralingsdosis of ergonomie het soms echter onmogelijk zulke uitgebreide data te verzamelen. In deze gevallen kan beelddata met een lagere dimensie of beelddata die schaars is in één van de dimensies verzameld worden.

We stellen in dit proefschrift voor om de oorspronkelijke dimensie van schaars verzamelde beelddata te reconstrueren door gebruik te maken van a priori kennis over de gewenste informatie. Deze informatie betreft de vorm of de vorm samen met de beweging van de vorm over de tijd. A priori kennis wordt geleerd aan de hand van een training set, die bestaat uit voorbeelden van mogelijke vormen en eventueel beweging van de vorm over de tijd. Met een dergelijke training set kan een model worden gemaakt dat bestaat uit het populatie gemiddelde en de modi van variatie binnen de populatie. Zulke modellen worden statistische vorm- of bewegingsmodellen genoemd.

We reconstrueerden de vorm uit schaarse data voor verschillende medische toepassingen. In hoofdstuk 2 stelden we een methode voor om de vorm van het dijbeen uit twee gekalibreerde röntgenfoto's te reconstrueren. Hiervoor minimaliseerden we de afstand tussen een statistisch vormmodel en de randen van het bot in de röntgenfoto's. Driedimensionale (3D) informatie over het botoppervlak is nodig voor het plannen van operaties en in recent ontwikkelde computer gestuurde operatiesystemen. Het botoppervlak kan bepaald worden uit CT scans, maar het reconstrueren van het oppervlak uit de röntgenfoto's is voor zowel de patiënt als voor het ziekenhuis voordelig, vanwege een lagere stralingsbelasting en lagere kosten van de behandeling. We testten onze methode met 10 in-vivo datasets en stelden een nauwkeurigheid van 1.68 mm vast. Een dergelijke nauwkeurigheid is vergelijkbaar met andere state-of-the-art reconstructiemethodes, maar onze methode vereist minder interactie van de gebruiker.

In hoofdstuk 3 werd de methode verder uitgebreid, om het reconstrueren en volgen van het dijbeen in twee gekalibreerde röntgensequenties (röntgenfilms) mogelijk te maken. Deze sequenties worden gebruikt voor het nauwkeurig afschatten van de beweging van de gewrichten, bijvoorbeeld voor het bestuderen van ligament-functie of

het beoordelen van de uitkomst van een operatie. In huidige bewegingsanalysemeth-
odes wordt de patiëntspecifieke vorm van het bot bepaald uit een vooraf gemaakt 3D
scan en frame voor frame uitgelijnd met de röntgenfilms. Onze methode maakt het
mogelijk de 3D scan te schrappen, waardoor kosten gespaard worden en de stralings-
belasting verlaagd kan worden. Verder maakten we het mogelijk om na eenmalige
interactie van de gebruiker de beweging van het bot automatisch te volgen. We
verwachten daarom dat de methode de bedieningstijd substantieel verlaagt. We li-
eten zien dat het reconstrueren van de vorm uit de hele sequentie een nauwkeuriger
resultaat gaf dan frame voor frame reconstructie, met een gemiddelde punt-naar-
oppervlakte afstand van 1.48 mm.

Het tweede gedeelte van dit proefschrift richt zich op de anatomie en beweging van
het hart. In hoofdstuk 4 stelden we een methode voor om, gegeven de vorm van het
hart, afgeleid uit een 3D CT angiografie (CTA) scan, patiënt specifieke hartbeweging
te voorspellen. Hiervoor bepaalden we aan de hand van een training set van 3D+tijd
hartsegmentaties en lineaire regressie de relatie tussen de vorm en beweging van het
hart. De voorspelde beweging kan helpen om de kransslagaderboom uit de 3D CTA
scan te projecteren op de röntgenfilms voor het verbeteren van de visuele feedback
tijdens minimaal invasieve percutane coronaire interventies (PCI).

Het afleiden van de vorm van het hart door het segmenteren van cardiale CTA
beelden gaat gepaard met een bepaalde mate van onzekerheid. Segmentaties en au-
tomatisch gevonden anatomische oriëntatiepunten zijn inderdaad nooit exact vanwege
bij voorbeeld ruis en eindige beeldresolutie van medische scans. In hoofdstuk 5 in-
troduceren we daarom een maximum-a-posteriori ridge regressie schema dat rekening
houdt met de onzekerheid in de segmentaties. In hoofdstuk 6 lieten we met een
grondige evaluatie zien dat deze methode beter werkt dan andere regressiemethodes
wanneer de onzekerheden in de test- en training set aanzienlijk verschillen.

Tot slot combineerden we de 2D/3D registratie uit hoofdstuk 2 en 3 met het
voorspellen van hartbeweging in hoofdstuk 7, en onderzochten we de toepasbaarheid
van statistische kransslagaderbewegingsmodellen voor het registreren van 3D CT
gebaseerde kransslagaders met intra-operatieve röntgen-sequenties. Deze registratie
is gewenst, omdat de kransslagaderboom uit preoperatieve 3D CTA scans informatie
bevat, zoals kalkdeeltjes of delen van de vaten, die niet zichtbaar zijn in de rönt-
genbeelden door totale verstoppingen. We verwachten dat het projecteren van deze
informatie op de intra-operatieve afbeeldingen de visuele feedback en de uitkomst van
de procedure voor patiënten met chronisch totale occlusies zal verbeteren. We stelden
vast dat statistische bewegingsmodellen ten minste zo goed presteerden als patiënt
specifieke bewegingsmodellen die afgeleid zijn uit 4D CTA scans. Dit resultaat is be-
moedigend, omdat het aantal 4D CTA acquisities vanwege dosisoverwegingen afneemt
ten gunste van 3D acquisities.

Publications

Journal papers

- **N. Baka**, M. de Bruijne, T. van Walsum, B.L. Kaptein, J.E. Giphart, M. Schaap, W.J. Niessen, and B.P.F. Lelieveldt. Statistical Shape Model Based Femur Kinematics from Biplane Fluoroscopy. *IEEE Transactions on Medical Imaging*, In Press

- C.T. Metz, **N. Baka**, H.A. Kirisli, M. Schaap, S. Klein, L. Neefjes, N.R.A. Mollet, B.P.F. Lelieveldt, M. de Bruijne, W.J. Niessen, and T. van Walsum. Regression-based cardiac motion prediction from single-phase CTA. *IEEE Transactions on Medical Imaging*, vol. 31, no. 6, pp. 1311-1325, 2012.

- **N. Baka**, B.L. Kaptein, M. de Bruijne, T. van Walsum, J.E. Giphart, W.J. Niessen, and B.P.F. Lelieveldt. 2D-3D shape reconstruction of the distal femur from stereo X-ray imaging using statistical shape models, *Medical Image Analysis*, vol. 15, no. 6, pp. 840-850, 2011.

- **N. Baka**, C.T. Metz, C. Schultz, L. Neefjes, R.J.van Geuns, B.P.F. Lelieveldt, W.J. Niessen, T. van Walsum, and M. de Bruijne. Statistical coronary motion models for 2D+t/3D registration of X-ray coronary angiography and CTA. *Submitted*

Conference papers

- **N. Baka**, C. Metz, M. Schaap, B. P. F. Lelieveldt, W. J. Niessen, and M. de Bruijne. Comparison of Shape Regression Methods under Landmark Position Uncertainty. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2011*, 2011, vol. 6892, pp. 434-441.

- **N. Baka**, M. de Bruijne, W. J. Niessen, J. H. C. Reiber, and B. P. F. Lelieveldt. Confidence of model based shape reconstruction from sparse data. *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2010, pp. 1077-1080.

- C.T. Metz*, **N. Baka***, H.A. Kirisli, M. Schaap, T. van Walsum, S. Klein, L. Neefjes, N.R.A. Mollet, B.P.F. Lelieveldt, M. de Bruijne and W.J. Niessen. Conditional shape models for cardiac motion estimation. *Medical Image Computing*

and Computer-Assisted Intervention - MICCAI 2010, Part I, LNCS 6361, 2010, pp. 452-459. * joint first authorship

- **N. Baka**, W. J. Niessen, T. van Walsum, L. Ferrarini, J. H. C. Reiber, and B. P. F. Lelieveldt. Correspondence free 3D statistical shape model fitting to sparse x-ray projections. *SPIE Medical Imaging*, 2010, pp. 76230D1 - 76230D9.

- **N. Baka**, J. Milles, E.A. Hendriks, A. Suinesiaputra, M. Jerosch-Herold, J. H. C. Reiber, and B.P.F. Lelieveldt. Segmentation of myocardial perfusion MR sequences with multi-band active appearance models driven by spatial and temporal features, *SPIE Medical Imaging*, 2008, pp. 691415-691415-10.

Conference abstracts

- **N. Baka**, M. de Bruijne, T. van Walsum, B.L. Kaptein, J.E. Giphart, M. Schaap, W.J. Niessen, and B.P.F. Lelieveldt. Fluoroscopic assessment of femoral kinematics using a statistical shape model. *20th meeting of the European Orthopedic Research Society, EORS 2012*, Accepted

- **N. Baka**, W.J. Niessen, B.L. Kaptein, T. van Walsum, L. Ferrarini, J. H. C. Reiber and B.P.F. Lelieveldt. Correspondence free 3D statistical shape model fitting to sparse X-ray projections. *Nederlandse Vereniging voor Patroonherkenning en Beeldverwerking NVPHBV, Fall meeting*, 2009

Workshop papers

- **N. Baka**, C.T. Metz, C. Schultz, L. Neefjes, R.J.van Geuns, B.P.F. Lelieveldt, W.J. Niessen, M. de Bruijne, and T. van Walsum. 3D+t/2D+t CTA-XA registration using population-based motion estimates. *MICCAI-Workshop on Computer Assisted Stenting, MICCAI-STENT'12*, Accepted

PhD portfolio

PhD period:	2007-2012
Departments:	Radiology, Medical informatics at the Erasmus MC
	Radiology at the Leiden University Medical Center
Research school:	ASCI

Courses

	By	Year
General academic skills		
Scientific English writing	Erasmus MC	2010
Course Giving a Presentation	Erasmus MC	2011
In depth courses		
Knowledge Driven Image Segmentation	ASCI	2007
Front-End Vision and Multiscale Image Analysis	ASCI	2009
Visualization and Virtual Reality	ASCI	2008
Summer school		
Molecular and medical image analysis and bio-informatics		2009

Presentations

International conferences	Type	Location	Year
Medical Image Computing and Computer-Assisted Intervention, MICCAI 2011	poster	Toronto	2011
IEEE International Symposium on Biomedical Imaging, ISBI	oral	Rotterdam	2010
Medical Image Computing and Computer-Assisted Intervention, MICCAI 2010	poster	Beijing	2010
SPIE medical imaging	oral	San Diego	2010
SPIE medical imaging	oral	San Diego	2008

Other meetings

Nederlandse Vereniging voor Patroonherkenning en Beeldverwerking, NVPHBV	oral	Leiden	2009
MXL european project meeting	oral	Leiden	2012
Research seminars Biomedical Imaging Group Rotterdam (4X)	oral	Rotterdam	yearly
Research seminars LKEB (4X)	oral	Leiden	yearly
Medical informatics lunch presentations (3X)	oral	Rotterdam	yearly
Radiology lunch presentation	oral	Rotterdam	2009

Attended seminars, meetings and workshops

	Location	Year
Mesh Processing in Medical Image Analysis, MICCAI 2011	Toronto	2011
Manifold Learning with Medical Images, MICCAI 2011	Toronto	2011
Statistical Atlases and Computational Models of the Heart, MICCAI 2010	Beijing	2010
Spatio-Temporal Image Analysis for Longitudinal and Time-Series Image Data, MICCAI 2010	Beijing	2010
Journal club	Rotterdam	weekly
Model based meeting	Rotterdam	biweekly
Image guided intervention meeting	Rotterdam	weekly

Teaching

	Year
Teaching assistant Introduction to Image Processing for Medical Students (Erasmus MC)	2010-2011
Teaching assistant Knowledge Driven Image Segmentation (ASCI)	2008, 2010, 2012

Other

	Year
Main organizer of the Medical Imaging Symposium for PhD students, MISP	2009
Reviewer for conferences: MICCAI	2011, 2012
Reviewer for journals: IEEE Transactions on Medical Imaging, Journal of Mathematical Imaging and Vision	2011-present

Acknowledgments

This thesis is the result of nearly five years of research at the Erasmus Medical Center and the Leiden University Medical Center. By being a "shared" phd student, I was fortunate to get insight into two groups, BIGR in Rotterdam, and LKEB in Leiden. I have learned much during these few years, both professionally and personally. In this chapter I would like to thank all who enabled this, taught me, supported me, and made my life enjoyable in this period.

First, I would like to thank all committee members for their time and effort to read my thesis and to come to its defense. Moreover, I would like to thank our secretaries Desiree, Tieneke and Petra for their help in organizing the defense.

Second, I would like to thank my two promoters, Prof. Wiro Niessen and Prof. Boudewijn Lelieveldt for letting me perform this research.

Dear Boudewijn, you were my master thesis supervisor, my co-promoter, and finally became my promoter. I started this PhD on your inspiration, as you encouraged me to consider doctoral positions after finishing university, and immediately offered me one on a familiar topic. I am very grateful for your trust and faith in me not only in the very early days, but also throughout the whole phd trajectory, and beyond.

Dear Wiro, your open leading style, energy, and love for the field made BIGR a unique research environment. Beside your theoretical knowledge and ability to manage people, I was most impressed by your overview of the developments in medical imaging. Sharing your thoughts on strategies and expected future trends was for me opening a refreshing, new view on the field.

My co-promoter, Marleen de Bruijne. Dear Marleen, I enjoyed the time under your supervision: you left me a lot of freedom with my research, yet I never felt alone on the way. Thank you for all the time you have spent on our discussions, and the patience you have listened to my sometimes crazy ideas.

Dear Theo, though officially not my co-promoter, you were involved in my project at several points. Thank you for enabling Chapter 7 of this thesis. I enjoyed working with your co-supervision.

I would also like to thank Erik Giphart for providing the data for chapters 2 and 3. Your enthusiasm on the results was always a great motivation for me.

Furthermore, I would like to thank my office mates on the 23rd floor, Coert, Michiel and later on Marius, for the enjoyable days. Michiel, you were always eager to help and think along. With your never ending stream of ideas you managed to keep me going every time I felt in a dead-end with my research. Coert and Marius, your personalities and humor made me really enjoy the CA days. Thanks for all

the conversations, and silly rides on the steps. Marius, special thanks for sharing the secret of Matlab and the old cluster. Coert, thank you for guiding me into the cardio-vascular field. Collaboration with you was very valuable to me, and resulted in two chapters of this thesis. You were always there to help with Mevislab/cluster issues/data, and basically everything. Special thanks for providing the latex source of your thesis as a starting point for writing mine, and for translating my summary into dutch. Furthermore, I would like to thank Hortense for inspiring me about being (at least a bit more) organized. Before meeting you I didn't quite appreciate this skill as it deserves. Hortense and Coert, I feel lucky to have you as my paranymphs.

Also on the social level I would like to thank all BIGR colleagues for the plenty of movie nights, drinks and BBQ-s. Hortense, Rahil, we made a great skate team! Stefan, Marcel and Dirk, camping with you in the Canadian wilderness was a great adventure. TT, it was fun getting lost with your guidance in Beijing (and find our ways again). Also thanks to Reinhard for the road-trip in the US, to Azadeh for the Iranian restaurant experience in San Diego, to Esben, for surviving the motorbike ride in Lipari, to Noemi, for the great time in Madrid, to Gerardo for your jokes so bad that they are good, and I could go on till I list everyone. I always had great fun, and hope to be able to keep in touch with you.

I also enjoyed my time in Leiden. I especially would like to thank new, old, actual and honorary KGB members for after work drinks, noise and silence at work (choose one), and good atmosphere. Vikas, Peter, Artem, Martin, Martijn, Ahmed, Ece, Trung, Ronald and Oleh, keep up the good spirits! Special thanks to Vikas and Artem to follow me up as MISP organizers, Avan and Andrei for making my first years more colorful, and Avan also for starting me off with ssm-s and matlab.

Since March I work at the department for Orthopedics in Leiden. I would like to thank Bart Kaptein and Edward Valstar to provide me with great freedom in my working hours, which made it easier to finish this manuscript next to working full-time. I also thank the researchers, especially at J9, for the good atmosphere, and Emiel for his contributions to the dutch summary.

Even though most of my time I have spent at work, my life would not have been complete without the support of friends outside of work. I would like to thank my friends in Hungary, who always made me feel at home upon arriving from the Netherlands. Lányok, bár néha hónapokig nem beszéltünk, minden tali olyan, mintha csak tegnap láttuk volna egymást. Azt mondják ilyen a jó barátság. Köszönöm.

Furthermore, I would like to thank all, who made my stay in the Netherlands so unique: my housemates at JVB, who showed me dutch student life, Matt and Solene for sharing a more quiet environment, all SoSalsa members, who always warmly received me despite that I often was not able to join any activity for months, and of course the Mexican guys, girls and relations, who were always ready to share a beer or tea on the weekends.

Természetesen mindez lehetetlen lett volna a család szerető támogatása nélkül. Anyu, Apu, köszönöm a bizamlatokat. Ti tettétek ezt lehetővé!

And finally, Diego, I am very thankful that I have met you during my PhD. Without your love, patience, and support, these years would not have been nearly as enjoyable and successfull as they were. Thank you!

Curriculum Vitae

Nóra Baka was born in 1983 in Kaposvár, Hungary. She received her masters degree in electrical engineering from Budapest University of Technology and Economics in 2007. At the end of her master studies she spent a year at Delft University of Technology (TUDelft) with Erasmus scholarship. During this period she conducted her master research at the Leiden University Medical Centre (LUMC) with close collaboration of TUDelft, which resulted in the thesis entitled: Knowledge Based Segmentation of MR Myocardial Perfusion Sequences. Subsequently, she joined the Biomedical Imaging Group (BIGR) at the Erasmus Medical Center, Rotterdam as a PhD student on the topic: Statistical models for anatomical shape and motion reconstruction from sparse image data. Her research was conducted in close collaboration with the Division of Image Processing (LKEB) at the LUMC, and has finally led to this thesis.

Currently, she holds a post-doctoral position at the department of orthopedic surgery at the LUMC. Her current research topic is statistical modeling of joints with a special focus on the shoulder and the knee.

Color Image Section

Chapter 1
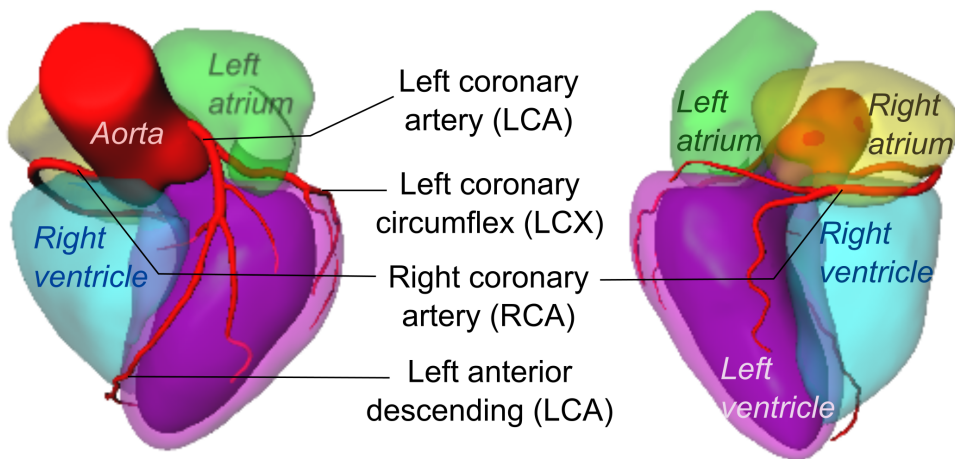
Figure 1.3 (see page 5): Illustration of the four heart chambers, the aorta, and the main coronary arteries. The coronary arteries directly branch off from the aorta, and provide blood to the cardiac muscles. Due to the transparency of the chambers both endo- and epi-cardium are visible for the left ventricle. The heart images were kindly provided by Hortense Kirisli.

Chapter 2
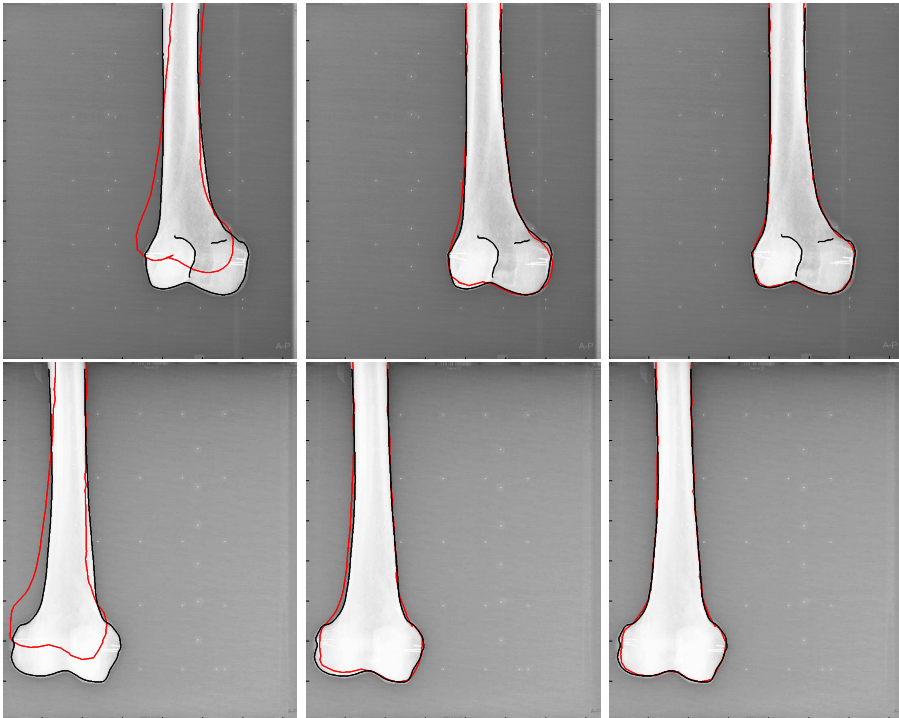
Figure 2.4 (see page 27): An example fit of Cadaver 1 in Anterior-Posterior pose. Canny edges are overlaid on the X-rays in black, red is the projection of the model. Top row: Left images, Bottom row: right images. From left to right: initialization $(P2S = 6.85mm)$, rigid fit with mean shape $(P2S = 2.3mm)$, non-rigid fit of the SSM $(P2S = 1.56mm)$.
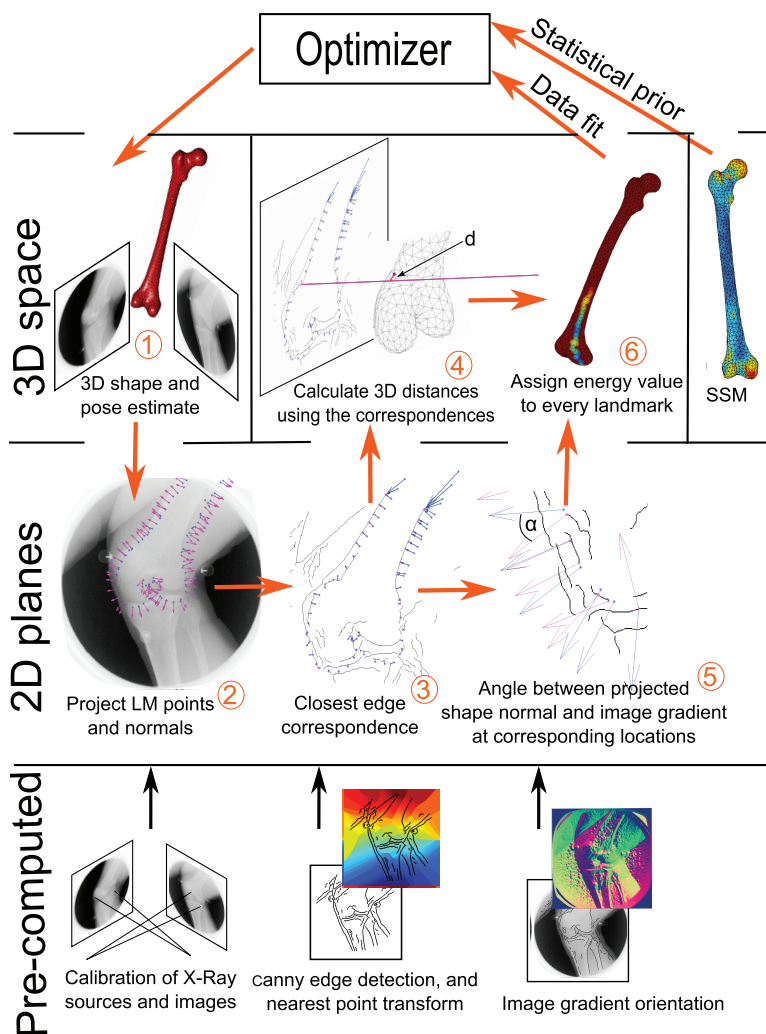
Figure 2.2 (see page 20): Method overview: The objective function is a combination of a shape prior term and a data fit term. The data fit term is calculated in six steps: 1) The silhouette landmark points of the current model are determined, and 2) projected to the plane; 3) The projected landmarks are matched with the closest edge point in the Canny edge map using a pre-computed nearest neighbor transform; 4) The 3D distance d between the silhouette landmarks and the ray connecting their corresponding 2D edge point with the X-ray source is defined; 5) The angle between the projected 3D surface normals at the silhouette landmarks and the gradient of the image at the corresponding edge positions are calculated; and 6) The matching score $\Theta_i(p_j)$ in Equation 2.5 is calculated for every landmark.

Figure 2.6 (see page 30): Top row: initialization Bottom row: matching scores (zero - dark blue; one - dark red) of the silhouette landmark points after SSM fitting. Ground truth shape is overlaid in green.
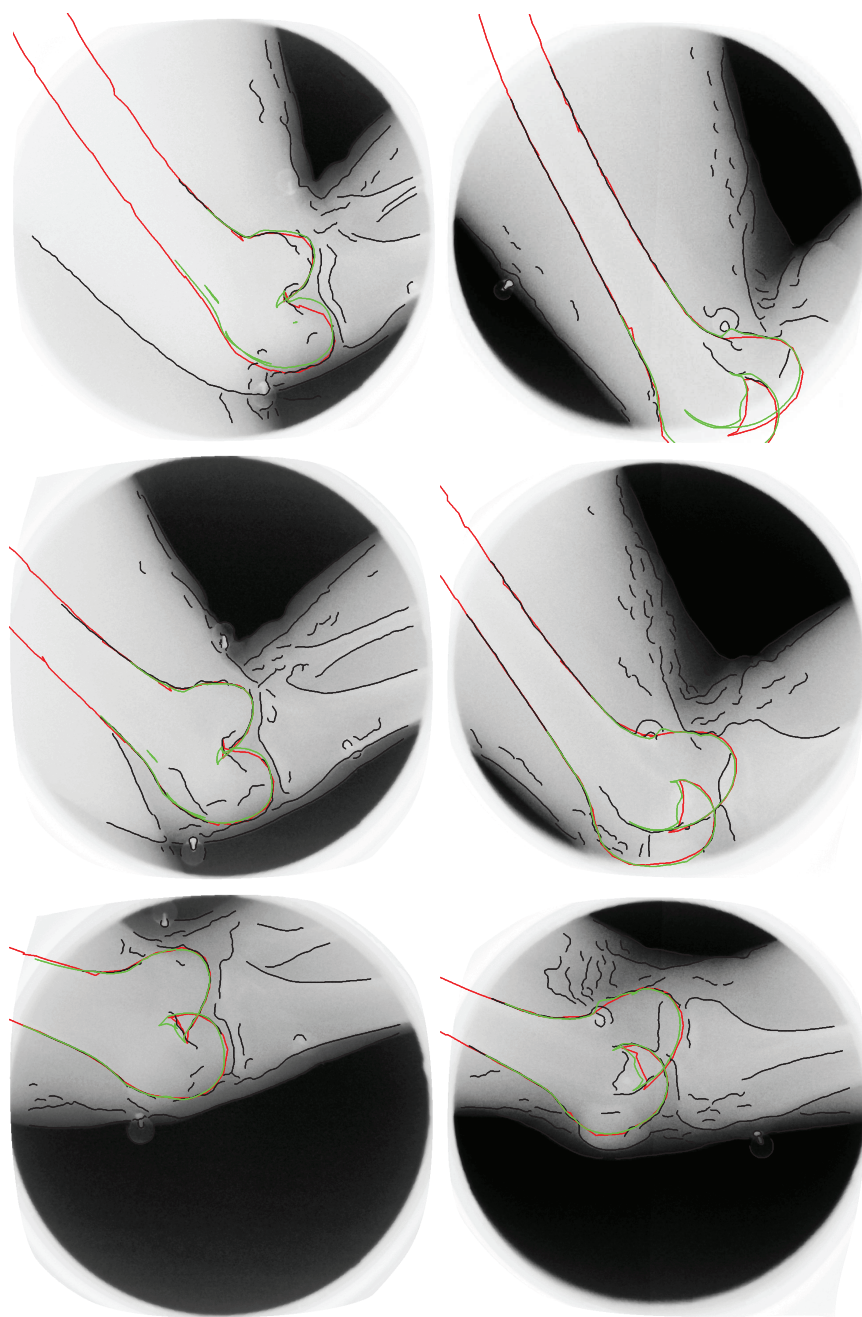
Figure 2.8 (see page 32): Examples of the SSM fit on the fluoroscopic soft landing sequence. Columns from left to right are frame 275, frame 307, and frame 331. Top row: Left images, Bottom row: Right images. Canny edges are overlaid on the X-ray with black. Red is the projection of the fitted shape, while green is the ground truth.
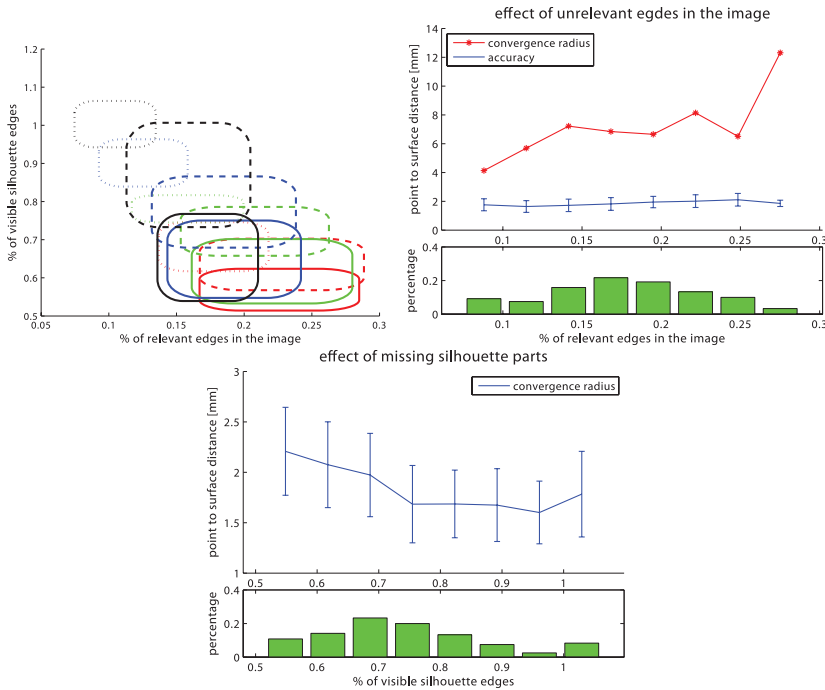
Figure 2.9 (see page 33): Effect of Canny parameters on the 10 fluoroscopic image-pairs. Top left: percentage of object edges, and percentage of visible silhouette edges for different Canny parameters. Colors code for the amount of Canny blurring: red $\sigma = 8$ pixels, green $\sigma = 6$ pixels, blue $\sigma = 4$ pixels, black $\sigma = 3$ pixels; and line style shows the threshold: 0.02 dotted, 0.04 dashed and 0.06 straight line. Top right: Convergence radius and accuracy of the 3D-2D non-rigid registration versus the percentage of object edges. Bottom: Convergence radius and accuracy of the 3D-2D non-rigid registration versus the percentage of visible silhouette edges. The histograms show the amount of test cases generating the plots.

Chapter 3
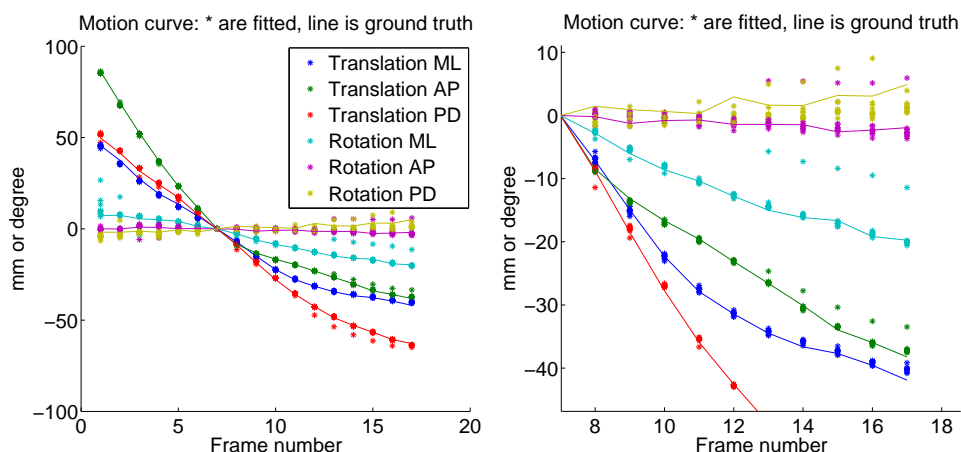
Figure 3.4 (see page 51): Detailed results of subject 7. Left: Motion curve with respect to frame 7. The solid line shows the ground truth motion, and stars represent the SSM fitting results of 10 runs with different starting poses. Right: a zoom-in to the time-points 8-17.
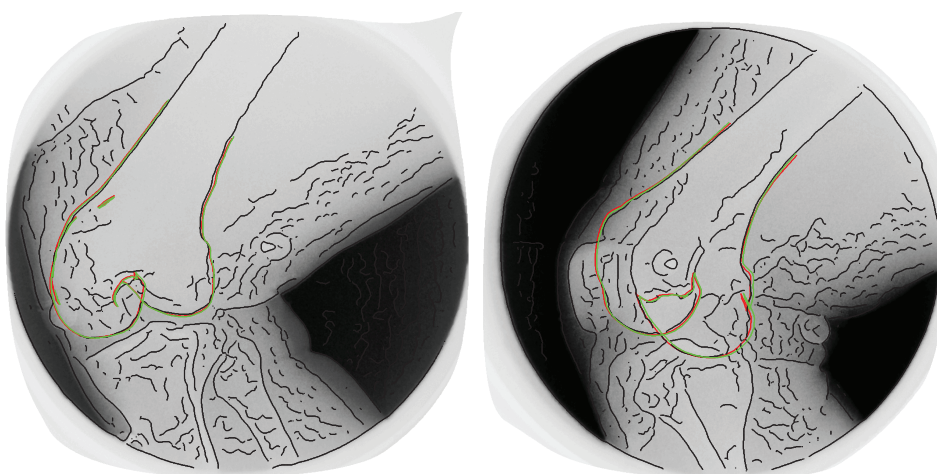
Figure 3.9 (see page 56): First and second planes of a category 1 time-point of subject 4 with the Canny edges and the fitted ground truth shape being overlaid. Red is the fitted ground truth shape, and green is ground truth shape in the reference standard pose. This time-point has PS distance of 0.65 mm.

Gradient orientation

KD-tree

for a fast closest edge-point selection

Edge locations

Fluoroscopic image

Edge features

Intensity

Edge length

Dynamic prior

Curvature

Prior knowledge of bone edge appearance

Expected pose of tracked object

Final edge relevance
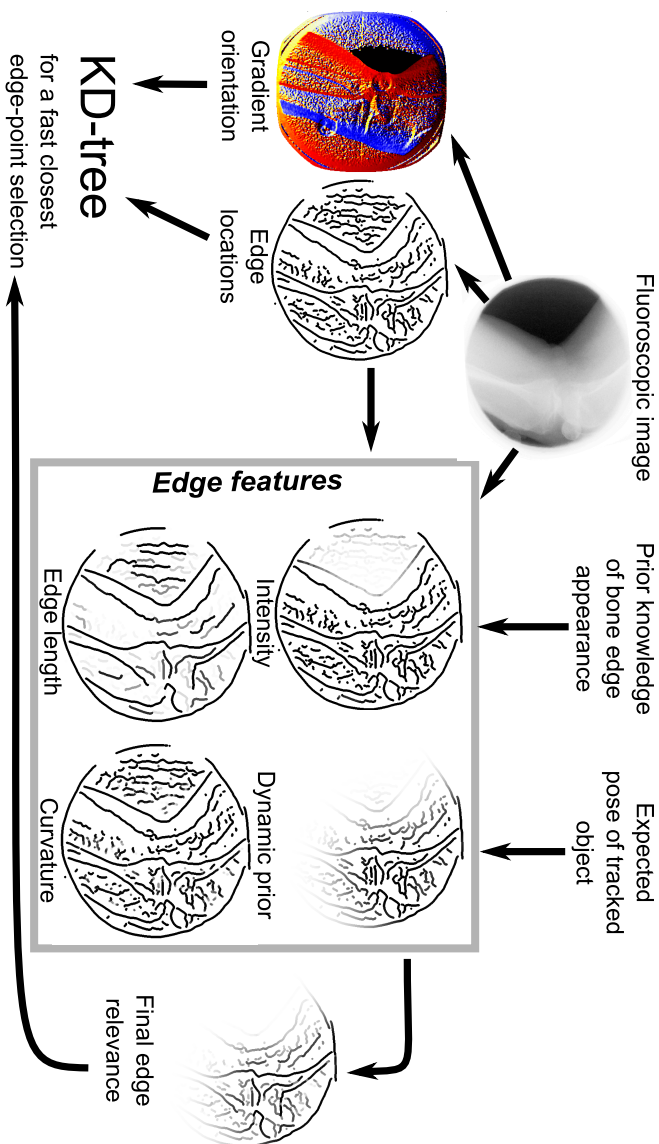
Figure 3.9 (see page 56): Correspondence building. Edge location, gradient orientation, and edge relevance are used in the correspondence determination. Edge relevance is a combination of edge features calculated from the original image, the Canny edge map and prior knowledge as explained in Section 3.2.4. Example images of one projection are shown with edges dilated for better visibility. The images are plotted with a linear colormap (1:black, 0:white).
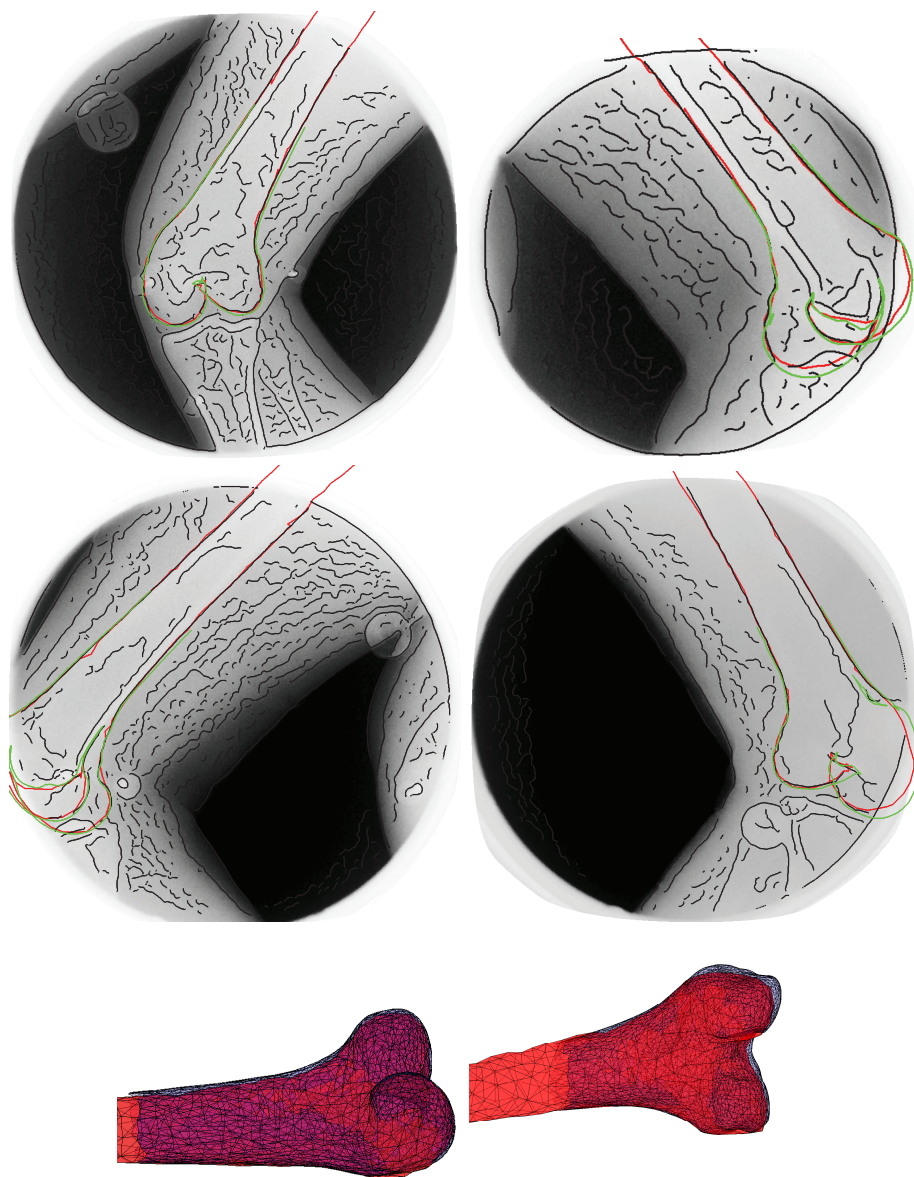
Figure 3.6 (see page 53): Two category 1 time-points of different subjects with overlayed contours of the fitted SSM (red) and the reference standard (green). From top to bottom: first plane, second plane, 3D surfaces of the reference standard (light blue) and the fitted SSM (red), rotated to see the differences. Left column shows a time-point with RMS PS distance of 1.20 mm (subject 7), and the right column shows a time-point with RMS PS distance of 1.74 mm (subject 3).
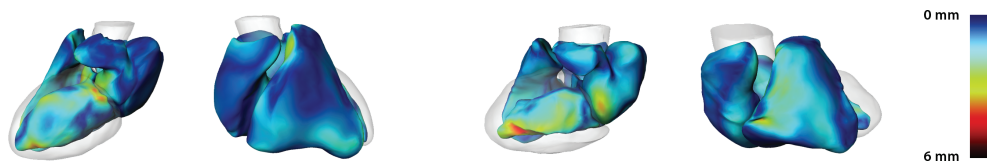
Chapter 4

Figure 4.3 (see page 67): Two views of end-systolic predicted surfaces for two randomly selected patients. Color coded are the point-to-surface distances from the predicted shape to the segmented shape. The Ao and epiLV are for visualization purposes not color coded.

Chapter 5
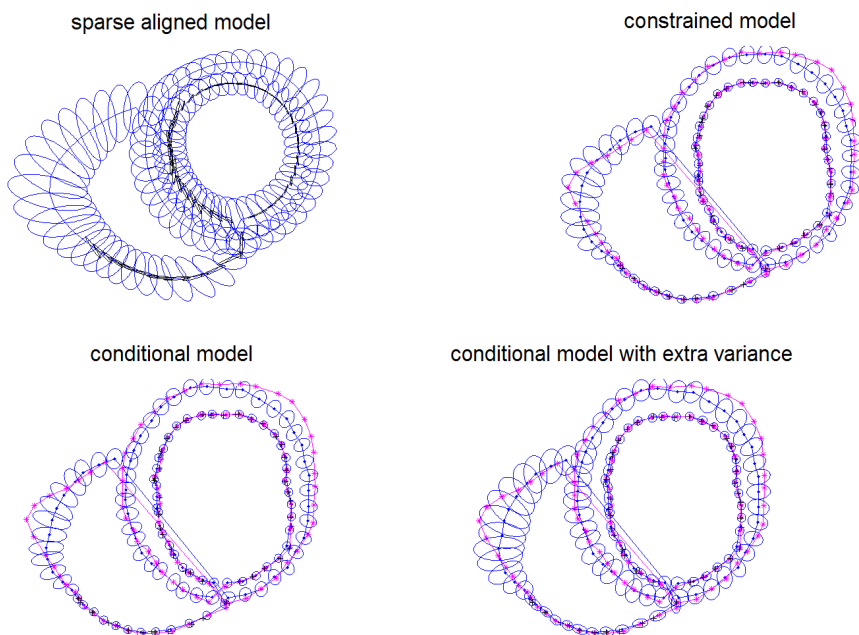
Figure 5.2 (see page 77): Confidence intervals of $\alpha = 0.95$ are indicated as blue ellipses (from left to right) on the sparse aligned model, the constrained model, the conditional model, and the conditional model with extra variance. The true shape is drawn magenta with landmarks indicated as *-s. Blue is the reconstruction, based on the black points +, which are noisy representations of the selected landmarks.

Chapter 6
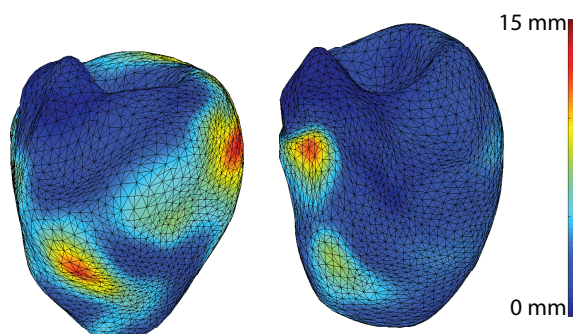
Figure 6.4 (see page 86): Two end-diastolic left ventricle shapes with color-coded landmark uncertainty [0-15 mm], derived from CTA data by multi-atlas segmentation.
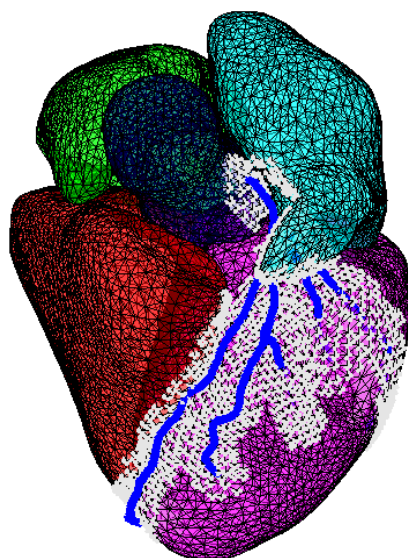
Chapter 7

Figure 7.1 (see page 94): End diastolic heart segmentation (four chambers and the aorta), displayed together with the coronary arteries (blue) and the selected nearby landmark points (white).
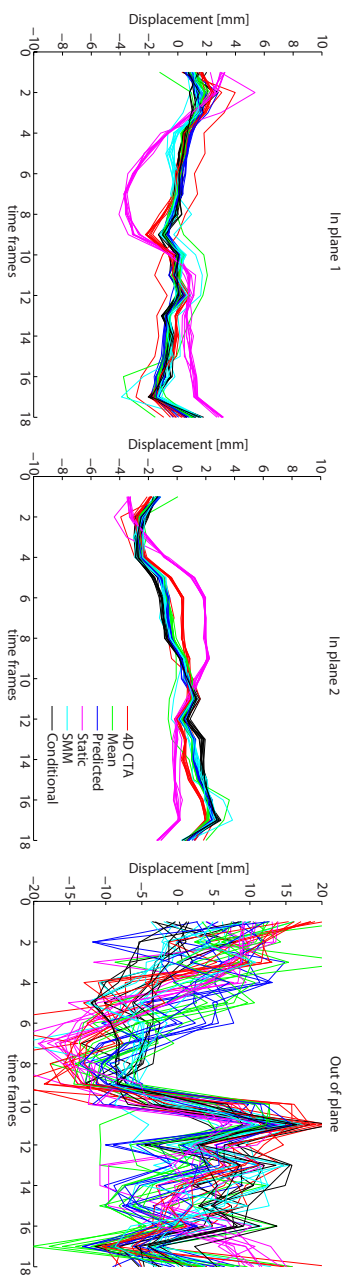
Figure 7.6 (see page 106): In plane (left, middle) and out of plane (right) translation for patient 2 through time when using the unconstrained respiratory motion model.
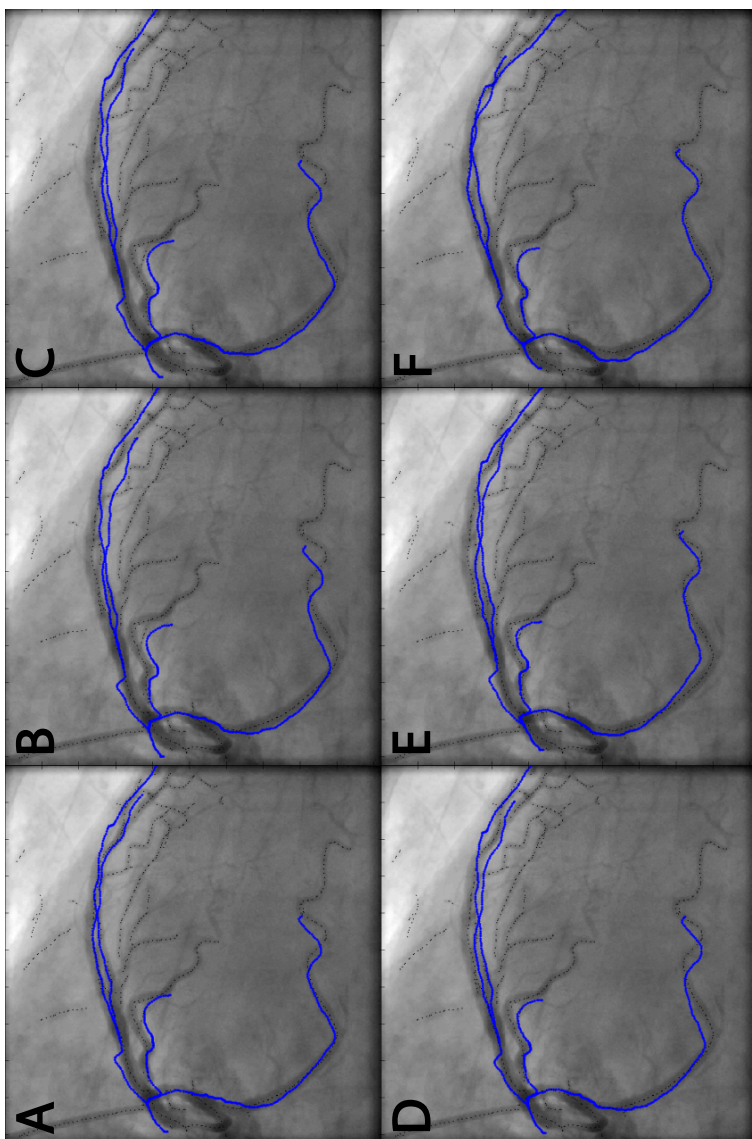
Figure 7.8 (see page 108): Example alignments with all coronary motion models and the quadratic respiratory model. The blue lines are the fitted coronary arteries overlayed on the corresponding systolic X-ray frame. Black dots show the subsampled centerlines used for fitting. A) 4D CTA motion, B) mean motion, C) predicted motion, D) static end-diastolic estimate, E) statistical motion model, F) conditional motion model.
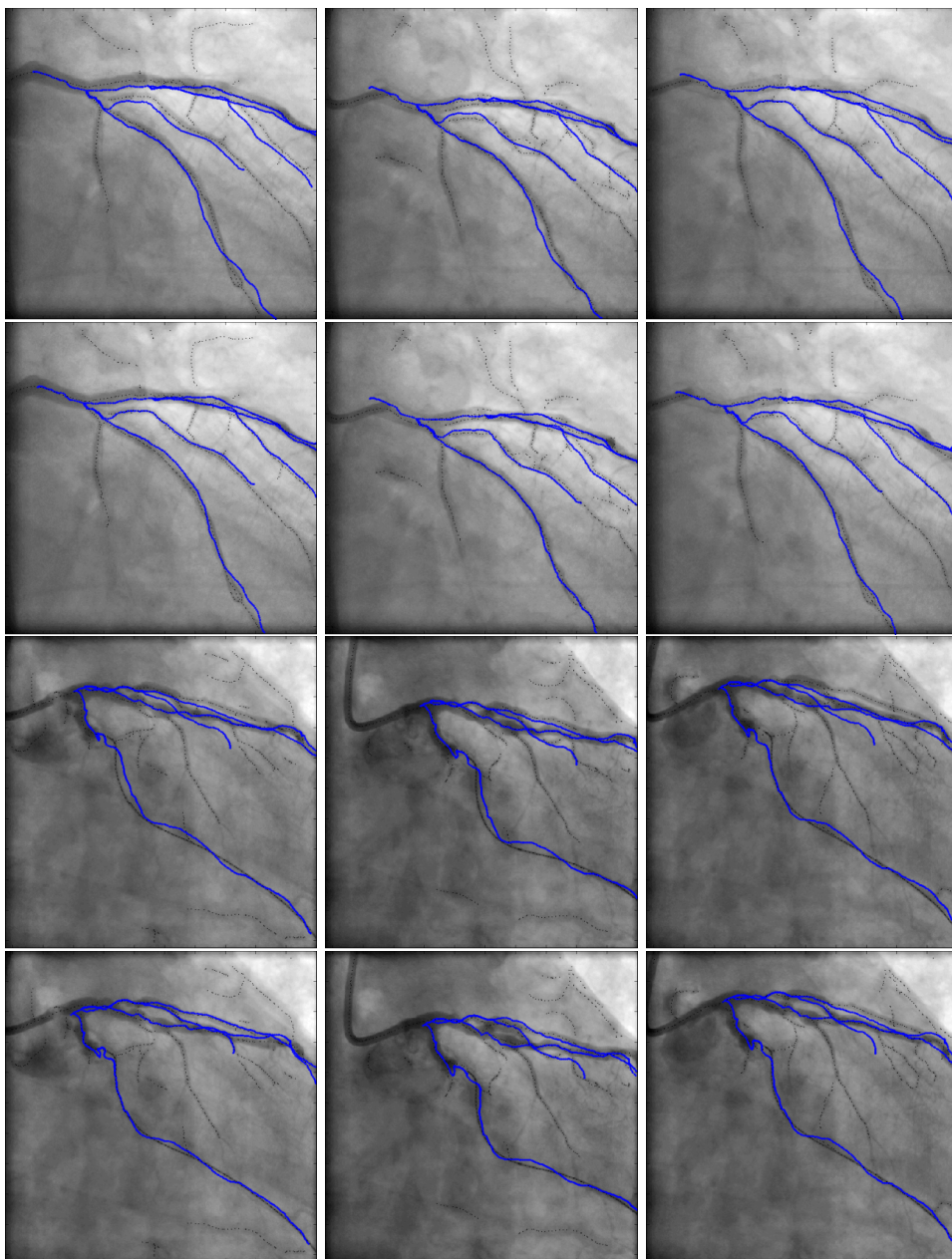
Figure 7.9 (see page 111): 2D+t/3D registrations of 3D vessels (blue) to XA data of two patients. Images are taken at 15%, 50%, and 85% of the cardiac cycle. First and third rows use the 4D CTA based coronary motion, the second and last rows use the SMM. A quadratic respiratory model was used in all cases.
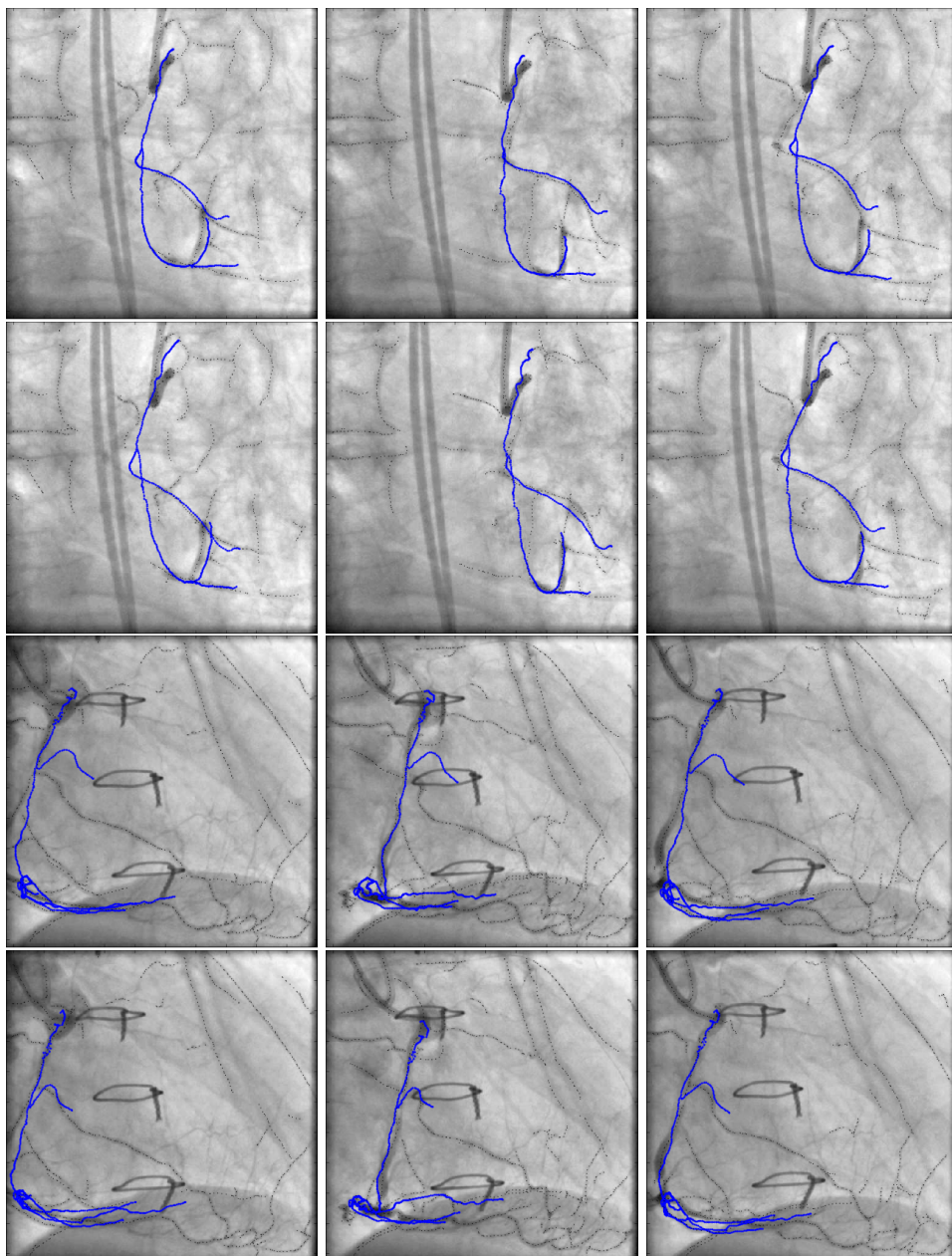
Figure 7.10 (see page 112): 2D+t/3D registration results (continued). Images are taken at 15%, 50%, and 85% of the cardiac cycle. First and third rows use the 4D CTA based coronary motion, the second and last rows use the SMM. A quadratic respiratory model was used in all cases.