2D-3D Registration of Knee Joint from Single Plane X-ray Fluoroscopy Using Nonlinear Shape Priors

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Jing Wu
May 2016

Copyright © 2016 by Jing Wu
All rights reserved.

DEDICATION

To my mother, Anfen Shao

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Mohamed Mahfouz for his valuable guidance and continued support of my Ph.D. research. I have learned a tremendous amount under his supervision.

I truly appreciate my committee members, Dr. Richard Komistek, Dr. William Hamel, Dr. Aly Fathy, Dr. Emam Abdel Fatah, for their precious time and invaluable suggestions on this dissertation. Especially I would like to express my deep gratitude to Dr. Hamel for his guidance and support at the early stage of my study for granting me the opportunity to the current research field which I am excited about.

I am very appreciative of the assistance extended to me by my colleagues, Dr. Michael Johnson, Dr. Michael Kuhn, Dr. Hatem El Dakhakhni, Dr. Sumesh Zingde, Dr. Nicholas Battaglia at the Center for Musculoskeletal Research (CMR). I would also extend my gratitude to Research Associate, Lyndsay Bowers, and Research Coordinator, Rebecca Robertson, for their support to provide me the opportunity to focus on my research.

I would like to express my deepest gratitude and thanks to my parents and my parents' in law for their continuous support and understanding through so many years of school. I would also offer my special thanks to my husband, Zongyao, for his endless support and understanding for my research work. His constant encouragement and support was, in the end, what made this dissertation possible.

ABSTRACT

Recent years have seen growing demands for computer-based information technology to help surgeons with surgical planning and guiding, kinematics analysis, computer-aided diagnosis (CADx), as well as prosthetics design. The key to this problem is the registration of two-dimensional (2D) intra-operative images to a three-dimensional (3D) model. This dissertation describes a novel non-rigid 2D/3D registration framework for reconstructing the 3D surface mesh model from a sequence of monoplane 2D fluoroscopic x-ray images based on nonlinear statistical shape model. The foundations for this framework are the following: 1) Feature Extraction, which extracts 2D contour from the X-ray fluoroscopy based on spectral clustering and active shape model (ASM); 2) Initialization, which estimates the 3D model's initial pose using a hybrid classifier integrating k-nearest neighbors (KNN) and support vector machine (SVM); 3) Optimization, which determines the 3D model's optimal pose and shape by maximizing the similarity measurey between the 2D X-ray fluoroscopy and the reconstructed 3D surface mesh model. The similarity measure is designed as a novel energy function including edge score, region score, homogeneity score, and multibody registration score. 4) 3D Shape Analysis, which represents the training dataset of 3D surface mesh models with nonlinear statistical shape model named kernel principal component analysis (KPCA). In conclusion, this work describes a novel, clinically relevant 2D/3D registration framework to provide a general approach that can be applied to solving medical image registration problems in a wide variety of modalities.

TABLE OF CONTENTS

CHAPTER I Introduction 1
1.1 Motivation 1
1.2 Method Overview 2
1.3 Computed Tomography (CT) 5
1.4 Statistical Shape Model 6
1.5 X-ray Fluoroscopy 7
1.6 Contributions 8
1.7 Organization 9
CHAPTER II Literature Review 12
2.1 Registration Classification 12
2.1.1 Degree of Freedom 12
2.1.2 Modality 13
2.1.1 Types of Registration 13
2.1.4 Dimensionality 14
2.2 Intrinsic Registration Methods 14
2.2.1 Library-based Registration 15
2.2.2 Feature-based Registration 16
2.2.3 Intensity-based Registration 17
2.2.4 Gradient-based Registration 19
CHAPTER III Method Overview 20
3.1 Problem Definition 20
3.2 Distortion Correction 22
3.3 Atlas Generation 24
3.4 Registration 25
3.4.1 Feature Extraction 25
3.4.2 Initialization 25
3.4.3 3D shape Analysis 27
3.4.3 Optimization 27
CHAPTER IV Feature Extraction 29
4.1 Introduction 29
4.2 Approach 31
4.2.1 Spectral Clustering 33
4.2.2 ASM based Segmentation 35
4.3 Experiments and Evaluation 41
4.4 Conclusions 47
CHAPTER V Initialization 48
5.1 Introduction 48
5.2 Approach 52
5.2.1 Problem Statement 52
5.2.2 Shape Extraction 53
5.2.3 Shape Representation 55
5.2.4 Estimate Out-of-plane Rotation by a Hybrid Classifier 60

5.2.5 Estimate In-plane Rotation 62
5.2.6 Pose Estimate Algorithm 63
5.3 Experiments and Results 65
5.3.1 Training Data 65
5.3.2 Experiments on Simulated Data 67
5.3.3 Experiments on X-ray Images 68
5.3.4 Feature-based 2D-3D Registration Experiments 70
5.3.6 Discussions 72
CHAPTER VI 3D Reonstruction from Monoplane Fluoroscopy 74
6.1 Introduction 74
6.2 Approach 77
6.2.1 Statistical Shape Model 78
6.2.1 Find the Closest Shape 83
6.2.2 Optimization 83
6.2.4 Energy Function 85
6.3 Experiments 90
6.3.1 Data 90
6.3.2 Experiment Design 92
6.3.3 Evaluation Scheme 92
6.4 Results 93
CHAPTER VII Knee Kinematics Analysis 99
7.1 Introduction 99
7.2 Approach 102
7.2.1 Energy Function 103
7.2.2 Edge Image 106
7.2.3 Sequence Optimization 107
7.3 Experiment and Results 109
CHAPTER VIII Conclusions 114
BIBLIOGRAPHY 117
VITA 129

LIST OF TABLES

Table 1. Compare the performance among original ASM, original ASM with relaxation, and spectral clustering based ASM. 46
Table 2: Classification accuracy of different kernel functions. The success rate is calculated with the error threshold set as 4 degree 61
Table 3: RMS error of synthetic data experiments with KNN-SVM classifier (90% patients for training and 10% patients for testing) 68
Table 4: Average Error of Experiments on x-ray images. Twenty x-ray images for each case 69
Table 5: Results of Initialized and Uninitialized registration 71
Table 6 Reconstruction accuracy 96
Table 7: Results of The Tracking With All Fits of The Five Test Sequences. Motion Errors Show The Mean Error with it Standard Deviation. The Unit for Translation Is in mm and for Rotation is in Degree 111

LIST OF FIGURES

Figure 1. System framework 3
Figure 2. Generation of 3d surface mesh model from CT scan. (a) CT scanner, (b) Volume data, (c) A slice of transversal image, (d) 3D surface mesh model (a copy from [2]) 6
Figure 3. Example of femur statistical shape model 7
Figure 4. Fluoroscopic imaging in the knee (a) fluoroscopic machine (b) an example of fluoroscopic x-ray image 8
Figure 5. Reconstruction of 3D model from a sequence of fluoroscopic images. 20
Figure 6. Schematic example of 2D-3D registration. 21
Figure 7. Fluoroscopic image of geometric calibration grid, (a) before distortion removal and (b) after distortion removal [74]. 23
Figure 8. 3D surface mesh models of femur and tibia 24
Figure 9. Flow chart of adding new bone to atlas. 25
Figure 10. Flow chart of adding new bone to atlas. 26
Figure 11. Data flow of the 3D reconstruction system. 28
Figure 12. 2D segmentation framework 32
Figure 13. Training landmark points in (a), aligned in (b) with $\overline{x}$ shown in blue. 38
Figure 14. Average cumulative variance as a function of the number of principal components. 39
Figure 15. Statistics of shape deformation. The dotted line represents the square root of eigenvalues sorted by decreasing size. The continuous line corresponds to the components of an individual model coefficient, which describes the deviation of the shape from the mean shape. 39
Figure 16. The gradient variation on the normal profile (search line) and the orientation of normal profile. 41
Figure 17. X-ray image enhanced by spectral clustering and Sobel operator. (a) original image; (b)-(h) first seven generalized eigenvectors of spectral clustering; (j) Maximum gradient response of the first seven eigenvector over orientations; (h) image enhanced by Sobel operator. The intensity of both spectral clustering and Sobel operator images are normalized to [0,1]. 42
Figure 18. Mean shape deformation using 1st, 2nd… 6st principal mode; (a)-(f) is the first to sixth principal mode. 43
Figure 19. An example of segmentation results by spectral clustering based ASM; (a) original image; (b) gradient search result; (c) result after ASM constraint; (d) result after relaxation; (f) final result. 44
Figure 20. (a)-(d) show qualitative results with ground truth in green and results from automatic segmentation in red. 45
Figure 21. Segmentation with shape prior on an image of femur occlusion from the neighboring bone: (a) original image, (b) segmentation with shape prior, (c) segmentation without shape prior. 46
Figure 22. Fluoroscopic imaging in the knee. 49
Figure 23. Flow chart of the initialization system. 54
Figure 24. Flow chart of 2D segmentation algorithm. 56

Figure 25. Segmentation result of femur and tibia in an X-ray image. 56
Figure 26. Reconstruction error of NEFD (a) RMS error of reconstructed femur contours with respect to the number of harmonics. (b) Reconstructed femur contours are in red with the number of harmonics from 1 to 50. The original contour is in blue. ...58
Figure 27. Compare the results of out-of-plane rotation by KNN and in-plane rotation by Procrustes Analysis. (a) Histogram of estimation of x and y axis rotation by KNN. (b) Histogram of estimation of z axis rotation by Procrustes analysis. 63
Figure 28. A diagram of SVMs set 64
Figure 29. 3D models of knee bone 66
Figure 30. Perspective effect by comparing two poses with the greatest variation in translation and fixed rotation during deep knee bending. (a) Projected shapes; (b) NEFD of the projected shapes.....67
Figure 31. Fluoroscopic X-ray images. 69
Figure 32. Overlaid of the estimated 3D model on the 2D X-ray image. 70
Figure 33. 3D reconstruction through nonlinear statistical shape model 78
Figure 34. Flow chart of finding closest shape 83
Figure 35. The two-stage sequence optimization. In the first stage the initial closest shape is rigidly aligned to the monoplane frames. k key poses are selected as input to the second stage optimization. The second stage alternates between shape and pose optimization until convergence occurs. 86
Figure 36. Edge score, (a) is the model edge image, (b) is the edge image extracted from fluoroscopic X-ray image 87
Figure 37. Region score. (a) is the model region image, (b) is the fluoro region image ..88
Figure 38. Energy function versus x and y axis translation. (a) edge score; (b) region score; (c) homogeneity score; (d) overall score. x axis is to the left, y axis is to the top, z axis is out of screen.....91
Figure 39. Compactness of the statistical models. Percentage of total shape variance versus the number of modes for statistical shape model using PCA and KPCA. 94
Figure 40. Generalization of the statistical models.....95
Figure 41. Specificity of the statistical models. (a) shows all the cases and (b) excludes the case of σ=100. 95
Figure 42. Femur and tibia 3D model edge overlay on the x-ray images in key poses during deep knee bending process 96
Figure 43. Femur surface distance map between reconstructed 3D model and the CT model.....97
Figure 44. Tibia surface distance map between reconstructed 3D model and the CT model 98
Figure 45. Method overview: The energy function is a linear combination of region score term E_r, an edge score term E_e, and a collision detection term E_c. The region score is defined by a normalized local cross correlation of model region image and the x-ray fluoroscopic image. The edge score is defined as a cross correlation of model edge image and edge image extracted from the original x-ray fluoroscopic image. The dynamic prior term uses information from the pose predicted in the previous frame as a regulation term. The multi-object registration term uses information from neighboring bones to avoid collision between bones and misalignment. The linear

combination of the above four terms forms the energy function, which is then minimized to get the final estimation of pose of the 3D model.....104
Figure 46. Edge image generation. (a) Fluoroscopic image; (b) gradient image; (c) edge image after thresholding; (d) edge image after removing small edges; (e) edge image after applying dynamic prior; (f) final edge image.....107
Figure 47. Energy function as a function of the number of iterations, on the same fluoroscopy image. 109
Figure 48. Monoplane fluoroscopic system for deep knee bending.....110
Figure 49. Model edge image overlay on X-ray images. 112
Figure 50. An example of a sample fitting result. (a) is 3D model overlay on the X-ray image; (b) is an overlay of projected 3D model edge image on the X-ray image..113

LIST OF ABBREVIATIONS AND SYMBOLS

2D	Two Dimensional
3D	Three Dimensional
CADX	Computer Aided Diagnosis
MRI	Magnetic Resonance Imaging
CT	Computed Tomography
PET	Positron Emission Tomography
fMRI	Functional MRI
TKR	Total Knee Replacement
OAI	Osteoarthritis Initiative
SSM	Statistical Shape Model
KNN	K-nearest neighbors
SVM	Support Vector Machine
PDF	Point Distribution Model
ASM	Active Shape Model
PCA	Principal Component Analysis
RBF	Radial Basis Function
KPCA	Kernel Principal Component Analysis
TM	Template Model
MCW	Mutual correspondence warping
RMSE	Root Mean Square Error
ICP	Iterative Closest Point
SVD	Singular Value Decomposition
OA	Osteoarthritis
DSC	Dice Similarity Coefficient
ICP	Iterative Closest Point
PA	Procrustes Analysis

CHAPTER I
INTRODUCTION

1.1 Motivation

Recent years have seen significant advances in medical imaging. Acquiring three-dimensional (3D) information about patient's anatomical structure has become an essential part of kinematics analysis, computer-aided diagnosis (CADx), treatment planning, surgical navigation, and prosthetics design in order to visualize complex geometries, understand the relationship between foreground and background objects, and perform measurement on the object of interest.

Pre-interventional 3D imaging is extremely valuable during orthopedic surgery for the precise reconstruction of complex structure and exact estimation of bone poses. Commonly used pre-interventional 3D images include Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography (PET), and Functional MRI (fMRI), which are three-dimensional modalities containing high-definition information about the anatomical structure. However, since these images are taken while the patient is in a static state, they may not completely reflect the actual situation. Moreover, since the generating three-dimensional images is time-consuming, it cannot be implemented during the intervention.

Intra-interventional images, on the other hand, provide up-to-date anatomical state during intervention. They are commonly two-dimensional X-ray, X-ray fluoroscopy, and portal images. Intra-interventional images are fast to generate; thus they can reflect changes caused by the procedure or to visualize a tool's location in real time. These images also

produce less radiation dose than pre-interventional images. However, 2D images contain much less information than the 3D modalities.

Therefore, fusing the information of 2D intra-interventional and 3D pre-interventional images is necessary. The fundamental step of the information fusion is a 2D/3D registration problem, which consists of finding a transformation from one modality to another such that all features that appear in one modality are aligned with their correspondences in the other. However, most of the existing 2D/3D registration systems are rigid registration [1-4], which requires pre-interventional images such as MRI and CT. These images involve time-consuming segmentation of the three-dimensional modalities and additional radiation dosage on the patient. Non-rigid 2D/3D registration is beneficial because 3D model is reconstructed from statistical shape model (SSM). Once the SSM is constructed, no CT or MR scan before intervention is required for each patient, thus lowering the radiation dose and acquisition costs.

Most of the non-rigid registration studies cited in the literature [5, 6] reconstruct the 3D model from two or more x-ray images, which may decrease the patient's freedom of motion in kinematical analysis. The goal of this work is to reconstruct the 3D anatomical structure from a sequence of monoplane x-ray fluoroscopic images so as to decrease the radiation dosage on the patient and to allow the patient's free motion during kinematical analysis.

1.2 Method Overview

This dissertation's goal is to develop a 3D reconstruction system for non-rigid registration of 3D model from a sequence of monoplane fluoroscopic X-ray images. The

method is generic and easily customizable to a variety of anatomical structures. The system framework is composed of four parts: initialization, feature extraction, optimization, and 3D shape analysis, as shown in Fig. 1.

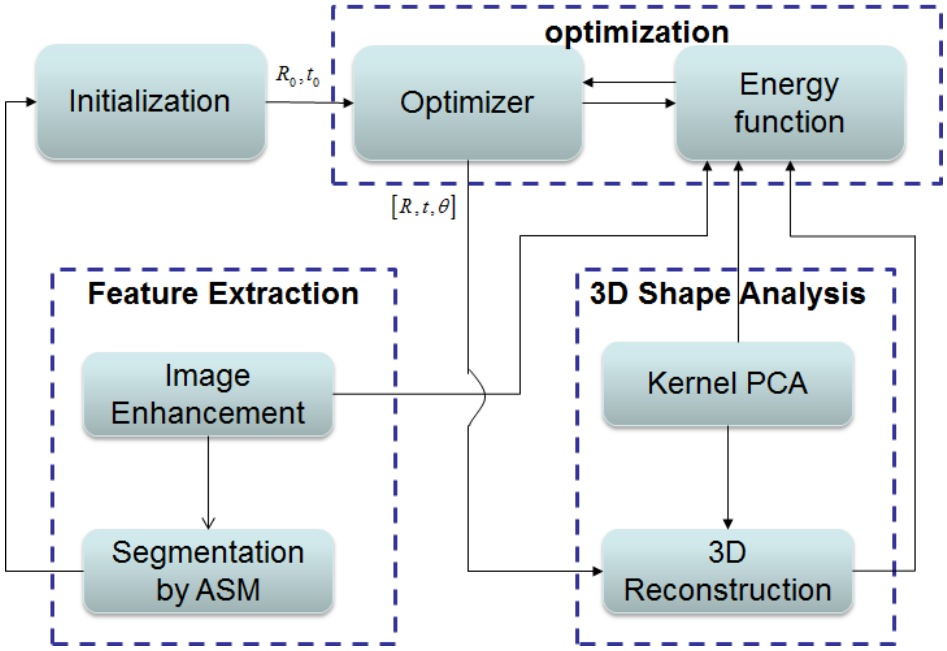


Figure 1. System framework

In the feature extraction module, a robust segmentation method accurately extracts the object of interest in 2D radiographs. The 2D radiograph is first enhanced by spectral clustering based on Eigen solution of an affinity matrix. A candidate contour is then generated by gradient search on the normal profile of each contour points. Constraint on the candidate contour is then performed with a statistical shape model, leading to the final segmentation result.

In the initialization module, the target anatomy's pose is estimated by finding the best match between the 3D model and a pre-computed library of shape templates. This

approach is quite flexible and is able to discriminate shapes with high variation. Contour of a knee bone or implant was first automatically extracted from a 2D radiograph as an output of the feature extraction module. Shape descriptors were calculated by normalized elliptical Fourier descriptors (NEFD) to represent the contour shape in a compact manner. The optimal pose was then determined by a hybrid classifier combining k-nearest neighbors (KNN) and a support vector machine (SVM).

In the 3D shape analysis module, a 3D statistical shape model is built from a dataset of surface mesh models segmented from CT images as training data based on kernel principal component analysis (KPCA). The training data is mapped into high dimensional feature space by a kernel matrix. Linear principal component analysis (PCA) is then performed in this high dimensional space to generate orthogonal basis of shape variations. Then, the 3D model is reconstructed from the selected shape parameters by a closed-form solution to optimize the pre-image approximation.

In the optimization module, a novel energy function includes edge, region, homogeneity, and multi-body registration scores to measure the similarity between the 3D model and the 2D radiograph. The energy function is minimized with a global optimization algorithm called pattern search (PS).

The proposed framework can register medical images from different modalities and it is flexible for either non-rigid or rigid registration by turning the shape parameters on or off.

The following are the data characteristics:

1. Statistical shape model (SSM) generated by a statistical shape atlas. The statistical shape atlas is composed of aligned 3D surface mesh models, which are segmented from CT images.
2. A sequence of X-ray fluoroscopies with 640×480 pixels, 8-bit gray-scale and pixel size of 0.225×0.225 mm^2, possibly including anatomy and surgical tools not present in the CT.

Details of the data characteristics are discussed in Section 1.3 – Section 1.5.

1.3 Computed Tomography (CT)

A CT scanner is used to preoperatively generate volume data. Fig. 2 (a) shows the CT scanner and Fig. 2(b) shows the volume data generated from CT scan.

The CT scanner consists of a translating bed, a scanner with a revolving X-ray source and a detector, and a monitor. To perform a scan, the scanner revolves around the imaged body on the bed and emits X-rays, which go through the body and hit X-ray detectors. By filtered back projection[1], a slice of transversal image of the volume data is generated; each slice is called a CT-slice (Fig. 2(c)). With the bed slowly moving through the scanner ring, a set of slices is obtained. From this set of CT slices, a 3D surface mesh model of the patient can be obtained by segmenting each CT slice (Fig. 2(d)). Pre-interventional images provide details about the patient's anatomy. However, since these images are taken while the patient is static, they may not completely reflect the surgical situation. Moreover, since generating the three-dimensional images is time-consuming, it cannot be implemented during intervention.

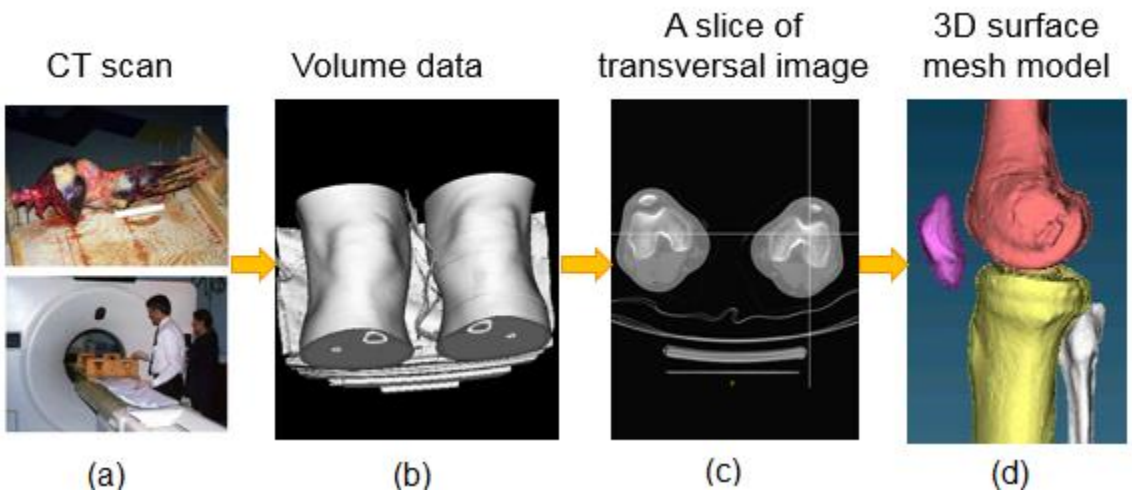


Figure 2. Generation of 3d surface mesh model from CT scan. (a) CT scanner, (b) Volume data, (c) A slice of transversal image, (d) 3D surface mesh model (a copy from [2])

1.4 Statistical Shape Model

Shapes extracted from CT can be represented by a statistical shape model (SSM). SSM represents the dominant shape variation in the training set. The training data are 3D surface mesh models that are segmented from CT image. The steps of building SSM are as follows: first, 3D surface mesh models extracted from CT are aligned to each other to build the 3D atlas. Kernel Principle Component Analysis (KPCA) [3] is then performed on the 3D atlas to extract the major modes of variation by first mapping the input data onto high dimensional feature space where a linear Principal Component Analysis (PCA) is performed on the data. The training results are the mean shape and the modes of variations captured by the first several principal components of the KPCA. A training process of femur SSM is shown in Fig. 3. SSM can represent the training data in a compact manner with KPCA.

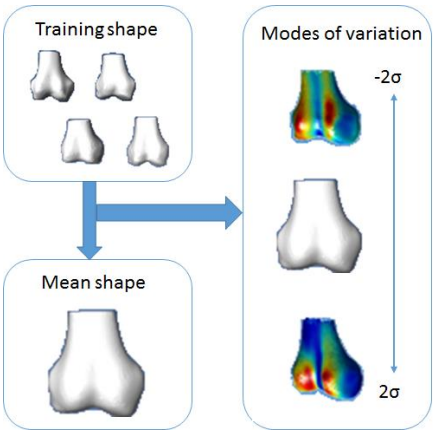


Figure 3. Example of femur statistical shape model

1.5 X-ray Fluoroscopy

Fluoroscopic X-ray images are a sequence of x-ray images, in which the bone appears much darker than the soft tissues surrounding them, allowing for direct observation and analysis of their movement under dynamic and weight-bearing conditions. Moreover, fluoroscopy is noninvasive and relatively low risk to the patient. A typical one-minute protocol gives the patient a radiation exposure on the order of 0.6 to 1.8 "rad equivalent man" (rem) [4]. However, 2D images contain much less information than the 3D modalities. And occlusion from foreign objects such as surgical tools and neighboring bones are common in x-ray fluoroscopic images. Fig. 4 shows the fluoroscopic machine in (a) and a fluoroscopic x-ray image captured in one frame of the video in (b) during the gait process.

Figure 4. Fluoroscopic imaging in the knee (a) fluoroscopic machine (b) an example of fluoroscopic x-ray image

1.6 Contributions

This research developed a non-rigid 2D/3D registration system, which is a novel general approach automatically registering medical images from different modalities. Theories from multiple disciplines (including computer vision, machine learning, signal processing and image processing) have been integrated into this research, which has produced the following extensive fundamental contributions in both theory and application:

1. Morphological analysis of tibia and femur using nonlinear statistical shape model as foundation for 3D reconstruction
2. Novel framework for 3D reconstruction of femur and tibia from single plane fluoroscopic X-ray images
    a) A novel closest shape estimation algorithm to estimate the initial shape of the 3D model

   b) A novel energy function combining edge, region, homogeneity, and multi-body registration score as a similarity measure
   c) A novel two-stage optimization framework to determine the shape and pose parameters of the 3D model
   d) Accuracy as good as or better than current state-of-the-art.
3. Video fitting of femur and tibia from single plane fluoroscopic X-ray images.
4. Novel segmentation algorithm for normal and OA knee in X-ray, fluoroscopic images
   a) Using Spectral Clustering for image enhancement, and statistical shape modeling for segmentation
   b) Accuracy shown to be as good as or better than current state-of-the-art
5. Novel initialization algorithm for estimating initial pose of 3D surface mesh model from single plane fluoroscopic X-ray image
   a) A novel two-stage pose estimation framework with a hybrid classifier
   b) Applied to pose estimation of femur and tibia
   c) Initialization significantly improved the 2D-3D registration accuracy compared with uninitialized registration

1.7 Organization

Chapter 2 reviews the related works in the following types of registration: library-based, feature-based, intensity-based, and gradient-based.

Chapter 3 presents an overview of the proposed system, which is composed of four parts: initialization, feature extraction, optimization, and 3D shape analysis. The

initialization module estimates the 3D model's initial pose with a hybrid classifier. The feature extraction module extracts the bone contour from 2D X-ray images based on SSM. In the optimization module, a novel energy function is designed including edge, region, homogeneity, and shape prior terms to measure the similarity between the 3D model and the 2D x-ray fluoroscopic image. In the 3D shape analysis module, a 3D statistical atlas model is built based on kernel principal component analysis (PCA).

Chapter 4 presents a feature-extraction method to extract the femur and tibia edges from 2D x-ray fluoroscopy. A spectral clustering-based method is used to enhance the image, based on which ASM is used as shape prior for segmentation.

Chapter 5 presents an initialization method that estimates the 3D model's initial pose with a hybrid classifier integrating k-nearest neighbor (KNN) and support vector machine (SVM). The 2D features are extracted from x-ray fluoroscopy with the method discussed in Chapter 4. Fourier descriptors are used to perform feature selection so that the shape outlines are encoded into a set of normalized descriptors as input to the classifier.

Chapter 6 presents 3D shape analysis by a nonlinear shape prior named kernel PCA. The 3D shape in the input space is mapped into a high dimensional feature space by a kernel matrix. PCA is performed in this high dimensional feature space to extract the most dominant shape variation in the training dataset. Then, the 3D model is reconstructed from the selected shape parameters by a closed-form solution to optimize the pre-image approximation.

Chapter 7 presents the application of 2D/3D registration in the multi-body video fitting process. A novel energy function is designed to integrate edge, region, and

alignment score to measure the similarity between the 3D model and the 2D x-ray fluoroscopic image.

Chapter 8 draws the conclusions.

CHAPTER II
LITERATURE REVIEW

The research objective is to design a robust, efficient and practical multimodality non-rigid 2D/3D registration system for medical images registration. As mentioned in the introduction, the system should be able to automatically register the medical images of different modalities with minimal human supervision. To achieve this objective, knowledge from several fields of study are required. In this chapter, the knowledge and results that are relevant to the research objective are reviewed. Section 2.1 presents several criteria for classifying registration methods. Section 2.2 elaborates on intrinsic registration methods, including library-based, feature-based, intensity-based, and gradient-based.

2.1 Registration Classification

Registration methods can be classified into several categories, according to degree of freedom, modality, types of registration, and dimensionality.

2.1.1 Degree of Freedom

Registration can be either rigid or non-rigid. Rigid registration determines only pose parameters and fixes the object's shape parameters. Non-rigid registration determines both pose and shape parameters. Rigid registration can be performed with single image [5]; while most of the non-rigid registrations in the literature requires at least two images for reconstruction [6-9]. Non-rigid registration is more challenging due to the extra number of parameters to be optimized in the registration process.

2.1.2 Modality

According to the image modalities involved, registration can be classified as quasi-intra-modality, multi-modality, and model-to-modality. The intra-interventional modality is an X-ray image for all types of registration. Different acquisition protocols and contrast agents may be used for x-ray images. They can be obtained at diagnostic energies (40-100 KeV) for high-contrast x-ray images or at treatment energies (2-20MeV) for low- contrast portal images. Other intra-interventional modalities include fluoroscopic X-ray, and ultrasound images. The pre-interventional modality varies among different registrations. In quasi-intra-modality registration, CT is the pre-interventional modality. Since CT and X-ray images are based on the same principle, they are the most closely related and thus the easiest to be registered among all types of registrations. In multi-modality registration, MRI is the pre-interventional modality, which is generated with different principle with X-ray image corresponding to specific patient. Multi-modality registration is in the middle in registration difficulty. In model-to-modality registration, 3D statistical model is the pre-interventional modality. Since the 3D statistical model takes no information about the specific patient, it is the least related and thus the hardest to be registered among all registration types.

2.1.1 Types of Registration

Registration can be either extrinsic when fiducial markers are used or intrinsic when registration relies only on images of anatomical structures. Extrinsic registration is very accurate but invasive, thus not always applicable. Intrinsic registration matches 2D images with projections of the 3D model. The various type of data further classify the intrinsic

registration into sub-classes: library-, feature-, intensity-, and gradient-based methods. The advantage of intrisic registration is its noninvasive, thus having wider applications.

2.1.4 Dimensionality

The dimension of the data in registration can be two dimensional (2D) for X-ray images or ultrasound and three dimensional (3D) for volumetric data for CT, MRI or CAD models. The dimensionality of extrinsic registration is 3D/3D, while that of intrinsic registration can be either 2D/3D for x-ray to CT registration or 3D/3D for MRI to CT registration.

2.2 Intrinsic Registration Methods

Intrinsic registration is technically much harder than the extrinsic one because it requires analyzing anatomical structures in intra-operative images. Moreover, foreign objects, such as surgical tools and implants that are not present in the pre-interventional images, may affect the registration quality. However, the non-invasive nature of the intrinsic registration makes it more practical than the extrinsic. The research in this area is active, and increasing progress has been reported since 1994 [10].

Because of the great potential in intrinsic registration, many methods have been developed for the 2D/3D registration of CT to x-ray images. Intrinsic registration is further classified into library-, feature-, intensity-, and gradient-based methods. The library-based method matches the 2D image with a pre-computed template library. The feature-based method matches geometric features extracted from different modality images by finding the transformation that minimizes the sum of corresponding feature distances[7, 10-13]. The intensity-based method directly matches the pixel intensity projected from 3D to 2D

images by minimizing the similarity measure between them [14-18]. This method does not need segmentation of pre-interventional 2D images, thus reducing the possible segmentation error in 2D images. However, generating a digitally reconstructed radiograph (DRR) is time consuming. To reduce the computational time, a gradient projection method is used to generate the original image's gradient and projects only the gradients above a set threshold [19-23].

2.2.1 Library-based Registration

Library-based registration determines the pose of the object of interest by matching its 2D contour extracted from the intra-interventional modalities with a pre-computed template library [24]. Each template is created by graphically rendering the object at a known pose. Since the accuracy of the library-based registration depends on the template resolution and fine resolution requires a large number of templates that are neither space nor time efficient, library-based registration usually serves as a coarse registration for initialization.

Two-dimensional contours of the object of interest can contain thousands of points, thus finding a compact representation for these contours is a critical step in the library based registration. This is a feature selection problem, which describes the shape in a compact manner while keeping the major information of the original shape. The methods of 2D shape description include global statistical approaches based on a method of moments[24, 25] and Fourier descriptors [26-30]. In Fourier descriptors, global shape features are captured by the first few low-frequency terms, while higher-frequency terms capture the shape's finer features.

2.2.2 Feature-based Registration

Feature-based registration determines the transformation that maximizes the similarity between features extracted from the pre-interventional images or a model and the corresponding features extracted from intra-interventional images. Features can be points such as landmarks and contours extracted from different imaging modalities. The feature-based method is fast because it uses only part of the information from different imaging modalities, but its accuracy depends on the segmentation quality. To reduce the influence of inaccurate segmentation, approaches for handling false geometrical position and outliers were developed [11, 31-37].

The simplest feature-based 2D/3D registrations are point-to-point registration of corresponding anatomical landmarks in both 2D and 3D images [11, 31, 38-40]. An operator usually locates landmarks; thus, the registration accuracy is heavily dependent on the operator's skills. This method often serves as a rough initialization after which a more accurate registration method is applied.

The problem of user-defined landmarks is avoided by curve-to-curve registration. When a projection strategy is applied, curves extracted from 2D and 3D images are matched by minimizing the sum of distances between corresponding curves [6, 7, 32, 41-43]. With the back-projection strategy, virtual rays from 2D points in the intra-interventional images are formed, and the distance between back-projected points in 3D space and the corresponding points on the 3D images are minimized [9, 44-48]. Zheng et al. proposed a hybrid approach combining the projection and reconstruction strategies. First, surface points on the 3D image are projected onto the 2D image to find the 2D correspondence. These corresponded 2D points are then back-projected to 3D space to

match the 3D surface points. Before calculating the distance between geometrical features, point correspondences must be established. Iterative closest point (ICP) is the most commonly used method to find the minimal Euclidian distance between points extracted from different modalities[49]. During registration, corresponding points are first determined by ICP; transformation is then implemented to update the new energy function for optimization. Alternatively, a pre-calculated distance map can be built to save calculation time [11, 31, 44, 50]. Since both of these methods are sensitive to outlier data, robust methods like adding directional correspondence between the curves as a weighing term in the energy function is used. The directional correspondence can be the angle between the tangents on the curves at corresponding points[8, 46], or between curves and surfaces[11, 13, 31, 34, 51]. Another method to suppress outlier points is to define outliers as points whose distance from the closest corresponding surface point after first pass of optimization is greater than the mean distance plus two standard deviations[52].

2.2.3 Intensity-based Registration

Intensity-based registration relies solely on the intensity information of pixels from 2D and voxels from 3D images. This method matches the intensity of one image modality with the intensity of the other by minimizing a measure of difference between them. In contrast to the feature-based method, intensity-based registration does not need segmentation; moreover, the correspondence in intensity-based registration is more explicit, thus making it more accurate than feature-based registration.

The registration of a CT volume to an X-ray image is estimated by iteratively minimizing the similarity measure between a simulated X-ray projection image from a

current transformation and X-ray image. Frequently used similarity measures include cross correlation [10], entropy [53], mutual information [54-57], gradient correlation [58], pattern intensity [18], and gradient difference [59]. Furthermore, a unifying framework for developing new similarity measures based on Markov random field modeling for different images was developed by Zheng [60]. From studies comparing different similarity measures [61-64], it can be concluded that similarity measures based on local-intensity correspondence produces more accurate and reliable results than the global correspondences. Pattern intensity and gradient difference are more suitable methods for lumbar vertebra registration [63].

The most used simulated x-ray projection images are digitally reconstructed radiographs (DRRs) by ray-casting a CT image. By averaging all the available information, the DRR-based method reduces the influence of outliers and is thus more robust than the feature-based method in the presence of foreign objects. However, the DRR-based method's drawback is its time-consuming generation process. Four main approaches to accelerate DRR computation are (1) fast-volume rendering [65-67]; (2) the hardware-based method [64]; (3) pre-computation of DRR [68-71]; and (4) segmenting 3D image before projection [5]. Rholfing [65] used progressive attenuation files and Birkfellner [66] introduced the wobbled splatting method to accelerate the DRR generation. The hardware-based method can be used, employing fast rendering withthe computer's graphic card to improve the registration speed up to 25 times [72]. An example of the pre-computed data structure is the Transgraph, introduced by LaRose [71] to accelerate the DRR generation. Segmentation of 3D images is performed to extract a surface model, while the registration framework remains the same as in the classic DRR-based registration [5].

2.2.4 Gradient-based Registration

Gradient-based methods were introduced by Tomasevic et al. [22] and Livyatan et al. [19] based on the fact that rays emanating from the X-ray source to the edge of X-ray images are tangent to the surface of anatomical structures. The gradient-projection relationship between gradients of attenuation coefficients and X-ray image-intensity gradients was built using a simplified model of X-ray generation, and thus the image intensity gradient at each point on the X-ray image can be back-projected toward the X-ray source to obtain a back-projected gradient at any location along the beam. This method requires establishing a 3D gradient vector filed from a CT or MR image before registration. The rigid 2D/3D registration is done by finding the best match between 3D gradient vectors and back-projected gradients.

A gradient projection-based method similar to the DRR method was introduced by Wein[23]. A 3D gradient image was calculated and 3D gradients above a user-set threshold were projected into 2D and matched with the 2D gradient image using the gradient correlation similarity measure. This method is 10 times faster than the intensity approach because it projects only the above-threshold gradients.

CHAPTER III

METHOD OVERVIEW

This chapter describes the 3D reconstruction framework and provides an overview of the system. Section 3.1 defines the problem to be solved; Section 3.2 describes the distortion correction; Section 3.3 discusses the atlas generation; and Section 3.4 provides an overview of the entire system.

3.1 Problem Definition

The goal of this research is to reconstruct 3D models from statistical shape model (SSM) using a sequence of single plane fluoroscopic x-ray images. It can be regarded as a non-rigid 2D/3D registration problem to determine a transformation of the 3D model so that it is aligned to the object of interest in the 2D image. The output is the reconstructed 3D models, which are aligned to the corresponding bones in a sequence of 2D fluoroscopic X-ray images, as shown in Fig. 5.

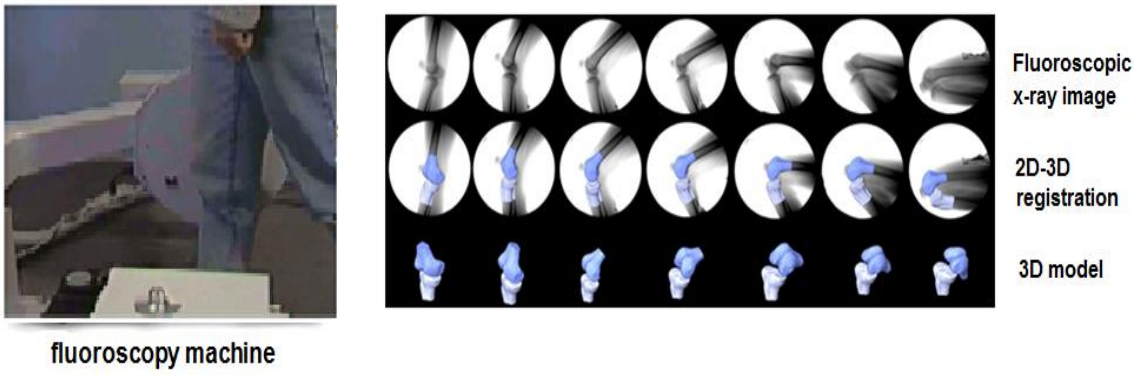


Figure 5. Reconstruction of 3D model from a sequence of fluoroscopic images.

The solution to the 2D/3D registration problem is twofold: first, to reconstruct the shape of the 3D model; and second, to determine the rigid transformation that relates the reconstructed 3D model to the 2D image with proper pose in the camera coordinate system. The relation among the X-ray source, the 3D model, and the imaging plane is shown in Fig. 6.

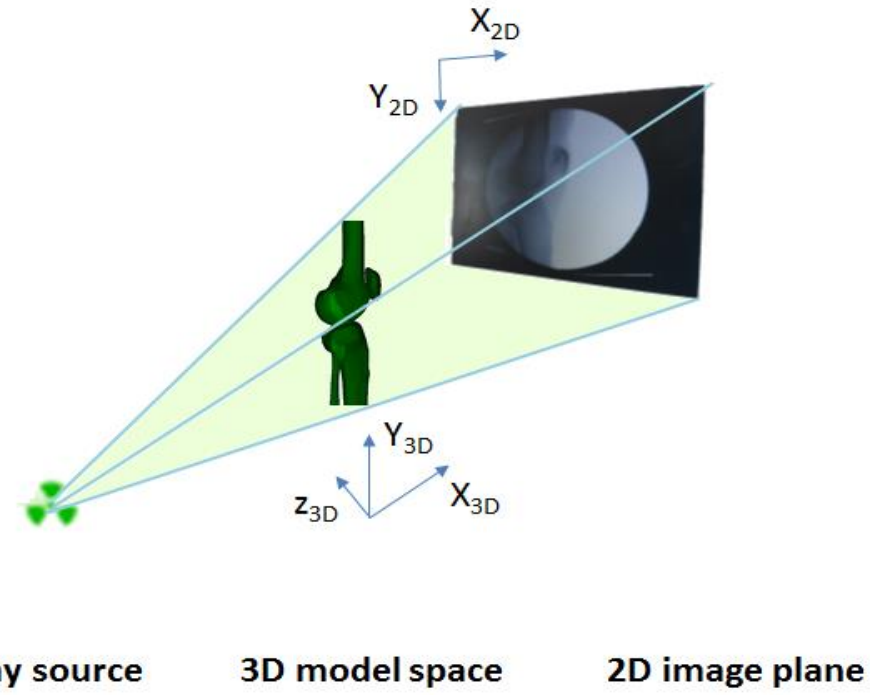


Figure 6. Schematic example of 2D-3D registration.

The fluoroscopic machine is modeled as a pinhole camera, with the camera's focal point at the X-ray source and the imaging plane at the image intensifier. Let $S \in \Re^3$ be the smooth surface of the object of interest, and denote $X = [x, y, z], X \in S$ to be the spatial coordinates measured with respect to the imaging camera's referential frame. Let $X_0 \in \Re^3$

and $S_0 \in \Re^3$ be the coordinates and surface in the 3D world, respectively. We can locate S in the camera reference frame via the transformation T such that $S = T(S_0)$ and the corresponding pointwise expression

$$X = T(X_0) = RX_0 + t \tag{3-1}$$

where R is a rotation matrix $R = R_\gamma * R_\beta * R_\alpha$, $R_\alpha = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}$,

$R_\beta = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}$, $R_\gamma = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and t is a translation vector, $t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$

.

X_0 in (3-1) is the reconstructed 3D model with the shape parameters, $\theta = \{b_1, \cdots, b_n\}$ via $X_0 = f(\theta)$ using kernel PCA, which will be discussed in Chapter 6.

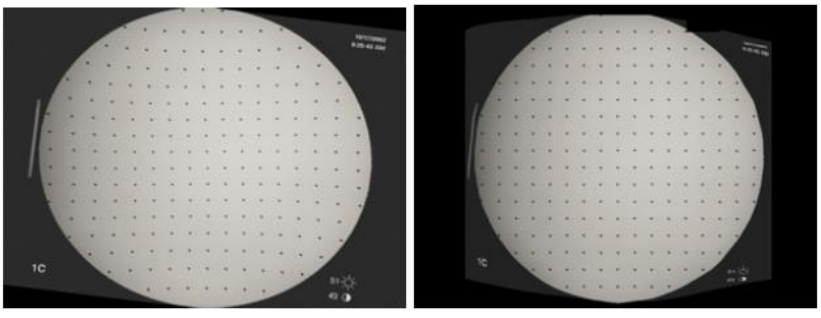
Therefore, two sets of parameters are to be determined to align the 3D model to the 2D x-ray image. One is the shape parameters $\theta = \{b_1, \cdots, b_n\}$. The other is the pose parameters $\theta = \{\alpha, \beta, \gamma, t_x, t_y, t_z\}$.

3.2 Distortion Correction

The image-creation process is necessary to be modeled before performing 3D reconstruction. Previous works have shown that the classic pin-hole camera model is a good approximation to this process, once geometric distortions presented in the fluoroscopic x-ray images have been removed [73]. The geometric distortions are caused

by two main factors: (1) the image-intensifier receptor screen is slightly curved, and (2) the Earth's surrounding magnetic fields and nearby instruments deflect the x-ray beam electrons.

Using a calibration target, we can estimate this distortion and remove it from subsequent images [5, 74]. The first step in our calibration procedure is to estimate any 2D image's geometric distortion. By taking an x-ray of a known rectangular grid of metal beads, we can estimate a 2D spatial transform for each small square sub-image that is bounded by four beads. Using standard techniques in geometric distortion removal [75], a local bilinear model is used to model the spatial mapping, as well as the gray level interpolation. Once the 2D distortion has been removed, the effective source-to-image plane distance (focal length) can be computed by a two-plane calibration grid with a known displacement between the planes. Fig. 7 illustrates the fluoroscopic image of geometric calibration grid before and after geometric distortion removal. We compute the bilinear transform for each set of four grid points that transforms the image positions of the beads in the left image to regularly spaced grid locations in the right. Clearly, the calibration procedure removes the pin-cushion distortion so that the grid points lie along straight lines.



(a) (b)

Figure 7. Fluoroscopic image of geometric calibration grid, (a) before distortion removal and (b) after distortion removal [74].

3.3 Atlas Generation

The training dataset consists of femur and tibia bones obtained from the William M. Bass Donated Skeletal Collection from the University of Tennessee Anthropology Department. These bones were scanned with a GE Light speed 16 Slice CT scanner. The resultant triangulated surface meshes were created in Mercury Amira®. An example of the 3D surface mesh model is shown in Fig. 8.

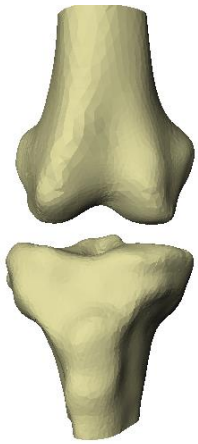


Figure 8. 3D surface mesh models of femur and tibia

Shape atlas was created by morphing the template model to match the training models. The training data was first aligned to the template, and then corresponding points between the training data and the template are found. The flow chart of adding new bone to the atlas is shown in Fig. 9. The detailed steps in creating the atlas are in [76].

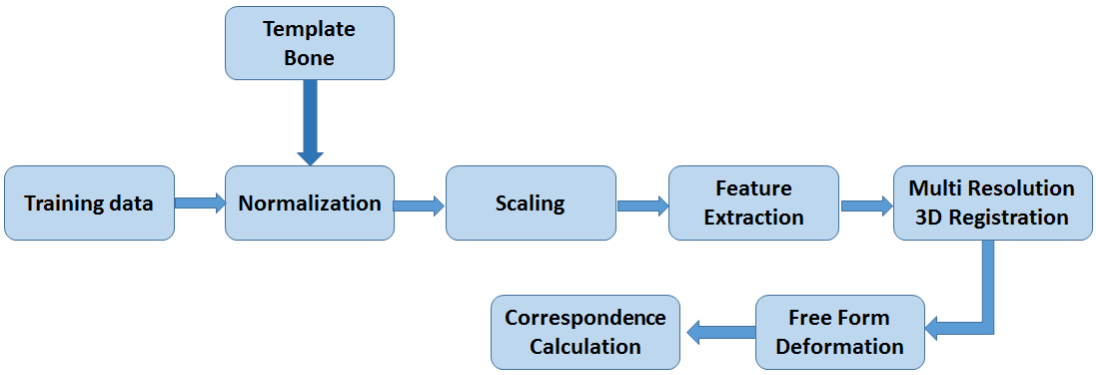


Figure 9. Flow chart of adding new bone to atlas.

3.4 Registration

This section presents an overview of registration algorithms, which are composed of four parts: feature extraction, initialization, 3D shape analysis, and optimization.

3.4.1 Feature Extraction

Feature extraction involves extracting 2D contour of the object of interest from the fluoroscopic X-ray images, which is a segmentation problem. The X-ray image is first enhanced by a spectral clustering-based method, and then segmented with the active shape model (ASM). The details about the feature-extraction algorithm are shown in Fig. 10.

3.4.2 Initialization

Initialization is to estimate the 3D model's initial pose from the first frame of the fluoroscopic x-ray image in the sequence. The initialization input is from the feature extraction where the 2D contour of the object of interest is extracted from the fluoroscopic

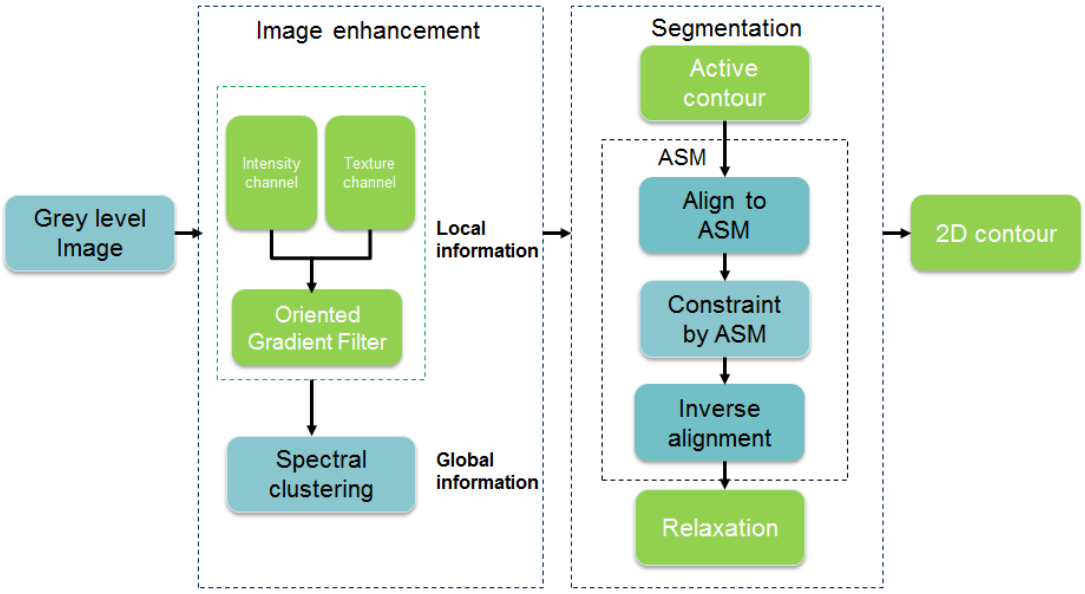


Figure 10. Flow chart of adding new bone to atlas.

x-ray image. A hybrid classifier integrating k-nearest neighbors (KNN) and support vector machine (SVM) is used to estimate the 2D contour's pose. The main steps are as follows:

1. Extract 2D contour from fluoroscopic x-ray image.
2. Encode the contour with normalized elliptical Fourier descriptors.
3. Estimate the out-of-plane rotation using KNN with the in-plane rotation normalized
4. Estimate the in-plane rotation with Procrustes analysis.
5. Re-estimate the out-of-plane rotation with SVM corresponding to the predicted in-plane rotation in Step 4.
6. Re-stimate the in-plane rotation with Procrustes analysis.

3.4.3 3D shape Analysis

A three-dimensional (3D) shape analysis is based on a nonlinear statistical shape model, namely kernel principal component analysis (KPCA). By projecting the training data onto high-dimensional kernel space, we can represent the shape of the 3D model by a vector of shape parameters. The 3D model is then reconstructed by a pre-image approximation.

3.4.3 Optimization

Optimization determines the 3D model's shape and pose parameters from a sequence of monoplane fluoroscopic X-ray images. Optimization is based on a novel energy function, which combines the edge, region, homogeneity, and multi-body registration score to measure the similarity between the 3D model and the 2D X-ray image. The hybrid energy function requires neither time-consuming DRR generation nor error-prone 2D segmentation.

An overview of the proposed method is shown in the flow chart in Fig. 11.

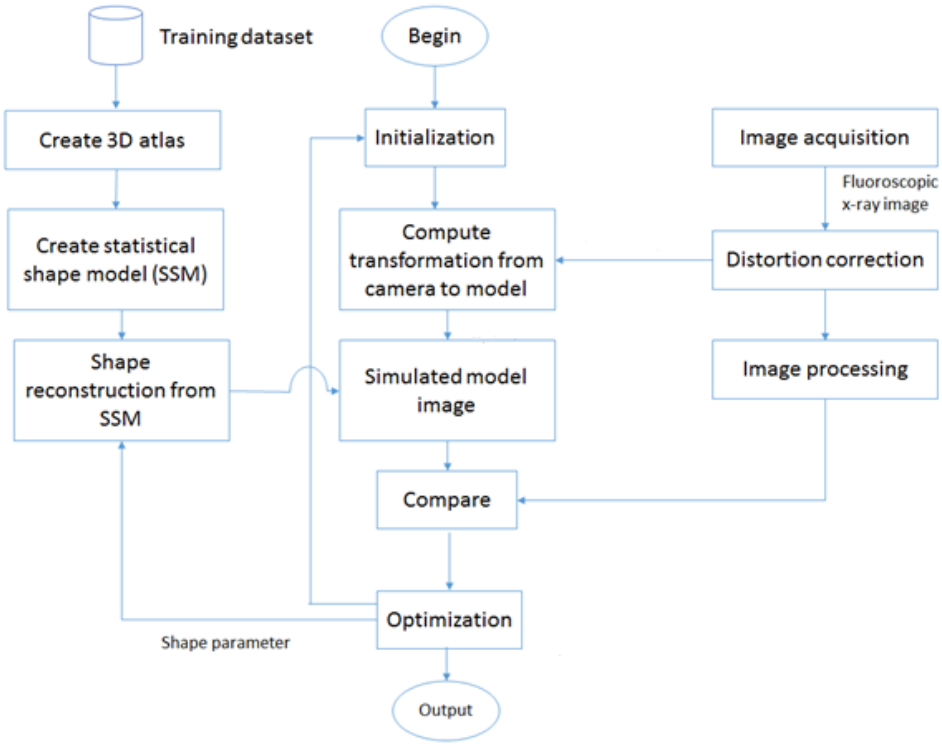


Figure 11. Data flow of the 3D reconstruction system.

CHAPTER IV
FEATURE EXTRACTION

Extraction of bone contours from X-ray radiographs plays an important role in joint space width (JSW) assessment, pre-operative planning, and kinematics analysis. We present a robust segmentation method to accurately extract the distal femur and proximal tibia in knee radiographs. The x-ray image is first enhanced by Spectral Clustering and then segmented with active shape model (ASM). In the following part, Section 4.1 is the introduction, Section 4.2 describes the detailed approach, Section 4.3 presents the experiments and segmentation results.

4.1 Introduction

Knee x-ray images segmentation has found wide applications in studying anatomical structure, kinematics analysis [4, 5, 77], assessment of joint loss of JSW, osteoarthritis (OA) diagnostics, joint replacement planning [78] and osteoporosis diagnostics of fracture detection and bone density measurement. Manual segmentation of these anatomical structures is time consuming and subjective. Our aim is to design an automatic segmentation method to extract the femur contour. Automatic segmentation of distal femur and proximal tibia from the x-ray images is challenging due to its complex structure. They may be influenced by noise, sampling artifacts, spatial aliasing, contrast, resolution or luminous intensities depending on the X-ray equipment so that the boundaries of the regions of interest become indistinct or disconnected. Inter-patient bone shape, size and deformation variations such as osteophytes adds complexity to the bone shape.

Moreover, overlapping with neighboring bones such as tibia and occlusion of implant or operational tools may cause loss of key features.

Previous work on automatic femur segmentation has primarily focused on active shape model (ASM) [79-82]. ASM was first developed in the mid-90's by Cootes [3] using a series of manually landmarked training images to generate a point distribution model, based on which principal component analysis was performed to generate a statistical shape model. There are several advantages in using ASM for segmentation: ASM is robust to artifacts and noise by the guidance of shape prior knowledge from training data of manually-annotated shape; have good performance for images with low signal-to-loise ratio (SNR); and is based on user defined forces. Behiels [79] added a regulation term in the cost function to impose smoothness of shape changes. Lindner [80] integrated random forest regression voting to ASM to segment proximal femur. Seise [82] used a double contour active shape model to segment tibia and femur from knee x-ray images.

A limitation of original ASM lies with using mahalanobis distance (MD) to find the object of interest. MD [83] is a statistical measurement to determine similarity between sets, which are assumed to have normal distribution. Normally the point with the minimum MD between its surrounding gray values and the mean gray values lies on the border of the object. To compute MD, a covariance matrix of the training gray value is constructed during the training phase and MD calculation uses the inverse of that covariance matrix. However, if the covariance matrix is sparse, the inverse matrix will be undefined, and consequently MD will be undefined. If the number of pixels sampled is greater than the number of training images, the covariance matrix will become sparse. Therefore, either having limited training data or attempting to sample a large number of pixels would prove

problematic. For active appearance model (AAM), original statistical appearance model assumes the appearance spaces to be Gaussian. However, in real cases the inter-patient bone shape varies significantly, especially for osteophytes bones. In such cases, a single mean mode is insufficient to capture the variations of appearance spaces with a single Gaussian distribution. Another limitation is the doubt on the training data to be sufficient to represent the testing case in AAM.

In this paper, we present a novel segmentation method in which spectral clustering is employed to consider the pixels with the similar appearance. Spectral clustering [84] is an unsupervised learning framework that relies on eigen-structures of a similarity matrix to represent the similarity between pixels. Spectral clustering partitions points into disjoint clusters with high intra-cluster similarity and high inter-cluster dissimilarity. From the similarity matrix, we derive a generalized eigenproblem and solve for a fixed number of eigenvectors to encode the contour information. Thereafter, we start searching from mean models to produce the candidates for segmentation. The statistical shape model derived from training dataset constraints the mean model in the feasible shape. This method captures large appearance variation, thus solved the limitations either by MD in ASM or limited training data in AAM.

4.2 Approach

The proposed method consists of two parts, spectral clustering to improve the image contrast and ASM for segmentation. The system framework is illustrated in Fig. 12.

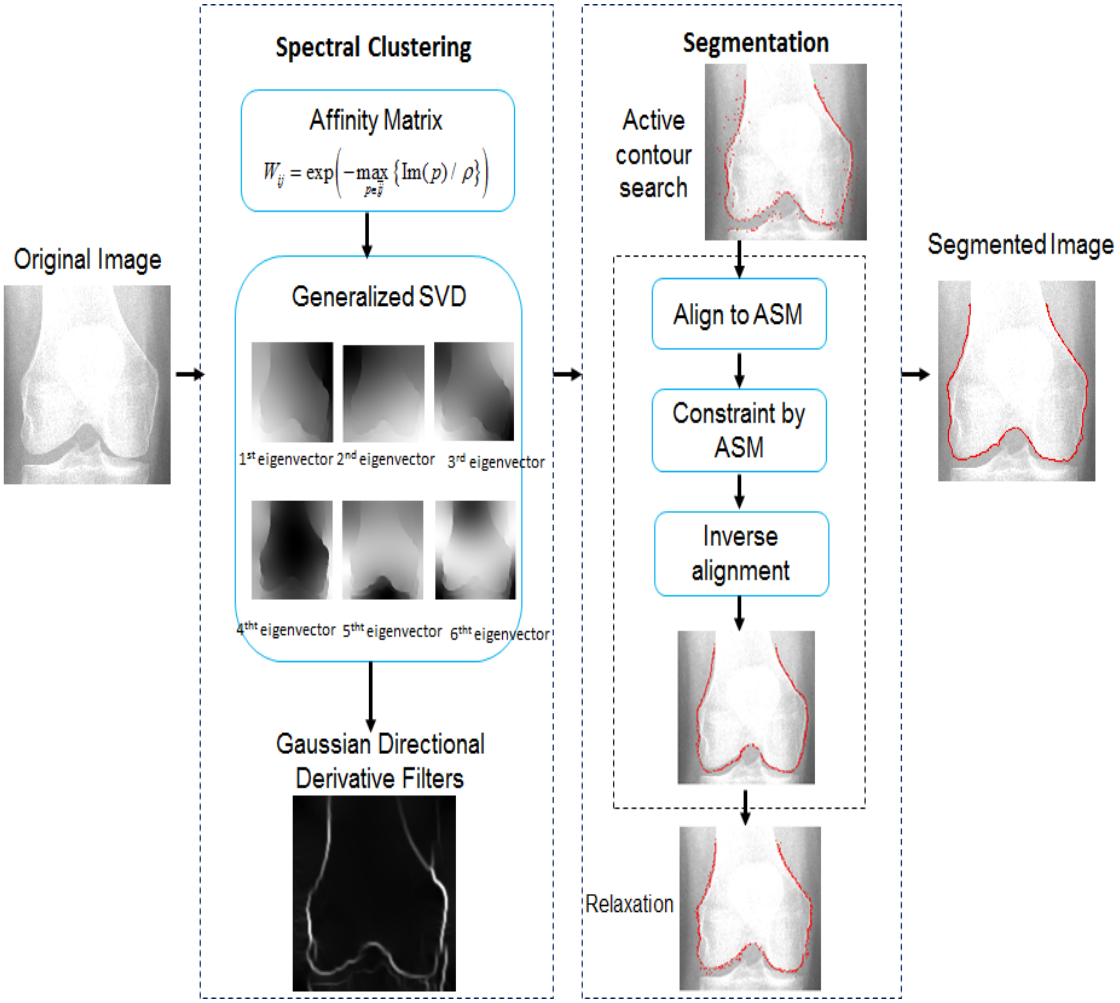


Figure 12. 2D segmentation framework

4.2.1 Spectral Clustering

The traditional appearance model (AAM) [85] is built from the intensity distribution of the region of interest. However, the intensities of the femur and tibia in the X-ray images may vary significantly with different x-ray equipment and X-ray intensity leading to an inaccurate appearance model. Therefore, to avoid the impact from inter-dataset intensity variations, we propose to determine the similarity between image pixels in an unsupervised learning framework of spectral clustering. Spectral clustering relies on the eigenvalues and eigenvectors of a similarity matrix to partition points into disjoint clusters with high intra-cluster similarity and high inter-cluster dissimilarity. We use spectral clustering in a similar manner to the Normalized Cuts approach [84] and use the eigenvectors obtained from spectral partitioning as weighted candidate for the following oriented segmentation rather than using a clustering algorithm such as K-means to create a direct partition of the image [86].

As an input to the spectral clustering, we construct a sparse symmetric affinity matrix *W* using the intervening contour cue [87], the maximal value of intensity (*Im*) along a line connecting two pixels. As is inspired from the intuition, Pixels *i* and *j* have a low affinity as a strong boundary separates them and a high affinity in the same region. The affinity matrix is formed by connecting all pixels *i* and *j* within a fixed radius *r* with affinity:

$$W_{ij} = \exp\left(-\max_{p\in\bar{ij}}\{\mathrm{Im}(p)/\rho\}\right) \tag{4-1}$$

where $\bar{ij}$ is the line segment connecting *i* and *j* and ρ is a constant.

A graph G = (V, E) can be separated into two disjoint sets, (A, B), where $A\bigcup B = V, A\bigcap B = \varnothing$. T based on the degree of dissimilarity between A and B, which is

computed by total weight of the edges that must be removed to separate the groups, called the "cut":

$$cut(A,B) = \sum_{i \in A,\, j \in B} W(i,j) \tag{2}$$

Shi and Malik [84] proposed a normalized cut to avoid cutting small sets of isolated nodes in the graph. It computes the cut cost a fraction of the total edge connections to all the nodes in the graph.

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)} \tag{4-3}$$

where $A \bigcup B = V$ and $assoc(A,V) = \sum_{u \in A, t \in V} w(u,t)$ is the total connection from nodes in A to all nodes in the graph and $assoc(B,V)$ is similarly defined. As a good segmentation corresponds to a partitioning scheme that separates all the nodes of a graph by cutting off the weakest links among them, i.e. minimize the cut value. In order to apply the optimal partition, we define $D_{ij} = \sum_j W_{ij}$ and x be an $N = |V|$ dimensional indicator vector, $x_i = 1$ if $x_i \in A$ and $x_i = -1$ otherwise. Thus the minimal cut can be found by computing:

$$\min_x Ncut(x) = \min_v \frac{v^T (D - W) v}{v^T D v} \tag{4-4}$$

with the condition $v(i) \in \left\{ 1, -\frac{\sum_{x_i > 0} d_i}{\sum_{x_i < 0} d_i} \right\}$ and $v^T D \mathbf{1} = 0$, where $v = (1+x) - \frac{\sum_{x_i > 0} d_i}{\sum_{x_i < 0} d_i}(1-x)$. As the Rayleigh quotient [88], (4-4) can be minimized by solving the generalized eigenvalue problem,

$$(D-W)\mathbf{v} = \lambda D\mathbf{v} \tag{4-5}$$

where $D_{ij} = \sum_j W_{ij}$, $\{\mathbf{v}_0, \mathbf{v}_1, ..., \mathbf{v}_n\}$ is to be solved as the generalized eigenvectors corresponding to the *n*+1 smallest eigenvalues $0 = \lambda_0 \le \lambda_1 \le ... \le \lambda_n$.

The generalized eigenvectors $\{\mathbf{v}_0, \mathbf{v}_1, ..., \mathbf{v}_n\}$ in (5) corresponding to the *n* smallest eigenvalues $0 = \lambda_0 \le \lambda_1 \le ... \le \lambda_n$ can be used for contour detection since the eigenvectors themselves carry contour information [86]. Thus, each eigenvector, $\mathbf{v}_i$, is treated as an image, and Gaussian directional derivative filters are applied at multiple orientations and combined with a weight of λ_i to provide the spectral information of the image:

$$s\,\mathrm{Im}(x, y, \theta)_i = \frac{1}{\sqrt{\lambda_i}} \nabla_\theta \mathbf{v}_i(x, y) \tag{4-6}$$

where the weighting by the eigenvalue λ_i is motivated by the physical interpretation of the generalized eigenvalue problem as a mass-spring system [89].

4.2.2 ASM based Segmentation

A. Training ASM

ASM represents the global shape constraints with the dominant shape variation in the training set. The training begins by manually selecting landmarks on the boundary in the femur x-ray images. The set of *N* training images is denoted as $S^t = \{C^\alpha \mid \alpha \in \{1, ..., N\}\}$, and each training image $C^\alpha \subset S^\alpha$ has a set of landmark $X^\alpha \subset C^\alpha$, where $X^\alpha = \{c_n^\alpha \mid n \in \{1, ..., N\}\}$, and where $c_n^\alpha = (x_n^\alpha, y_n^\alpha)$ denotes the coordinates of the n^{th} landmark point in C^α.

Two steps of alignment are necessary before building the shape model. First alignment is to find the corresponding points between each contours using statistical atlases, where a training contour with known point distribution is deformed to match the training set [90]. First, the centroids of the training contour and the new contour are aligned and the training contour is pre-scaled to match the bounding box dimensions of the new contour. Second, a rigid alignment of the training contour to the new contour is performed using a standard vertex-to-vertex iterative closest point (ICP) algorithm [91]. Third, a general affine transformation is performed to align the training contour to the new contour using 12 DOF (rotations, translations, scaling, and shear). After the affine transformation step, the training and new contour have reached the limits of linear transformation, but local portions of the contours still remain significantly distant. The goal of final matching is to create new points on the new contour, which will have similar local spatial characteristics as the training one. Second alignment is to align the training contour by translation, rotation, and scaling with least square estimation [92, 93]. The detailed algorithm for alignment is shown is Algorithm 1. Fig. 13 (a) and 1(b) shows the training landmark points before and after alignment for the representative knee X-ray images.

Once the training shapes have been aligned, the mean shape $\bar{X} = \left\{ \bar{c}_n \mid n \in \{1,...,N\} \right\}$ where $\bar{c}_n = \left(\bar{x}_n, \bar{y}_n \right)$ and $\bar{x}_n = \frac{1}{K} \sum_\alpha x_m^\alpha$ and $\bar{y}_n = \frac{1}{K} \sum_\alpha y_m^\alpha$. Then, PCA is used to the aligned shapes [3], so that any valid femur shape x can be represented as

$$\mathbf{x} = T\left(\bar{\mathbf{x}} + \mathbf{\Phi}_s \mathbf{b} \right) \tag{4-7}$$

where T is a linear transformation mapping, $\bar{\mathbf{x}}$ is the mean shape, and $\mathbf{\Phi}_s = (\phi_1, \phi_2, ..., \phi_m)$ is a matrix of eigenvectors corresponding to the *m* largest eigenvalues λ_i derived from the covariance matrix, $S = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$. *m* is chosen to be the smallest number satisfying $\frac{\sum_{i=1}^{m}\lambda_i}{\sum_{i=1}^{n}\lambda_i} > 0.99$, representing more than 99% of the variance of the training set.

These eigenvectors ϕ_i represent orthogonal basis of linear deformation modes that describe how points tend to move together as the shape varies. The corresponding eigenvalue λ_i is equal to the variance described by each linear deformation mode. **b** is the shape model parameters, computed by $\mathbf{b} = \mathbf{\Phi}_s^T(\mathbf{x} - \bar{\mathbf{x}})$, when fitting the model to a set of points, the values of **b** are constrained to lie within the range $\pm 3\sqrt{\lambda_i}$ [30].

Algorithm 1 Normalization of training data

Input: Training contours (x_1,y_1,……x_n,y_n)

Output: Aligned training contours (x'_1,y'_1,……x'_n,y'_n)

1: Translate contours to (0, 0)

2: Set the first contour x(0) as initial estimate of mean model and scale it so that $|x(0)| = 1$

3: Align all the remaining shape to the mean shape by translation, rotation, and scaling with least square estimation

4: Re-calculate the mean model from the aligned contours

5: Apply constraints on the current estimate of the mean model by aligning it with previous mean model and scaling so that $|\bar{x}| = 1$

6: Calculate new mean model

7: If $\frac{\sqrt{\sum_i (\bar{x}_{i,current} - \bar{x}_{i,previous})^2}}{\sqrt{\sum_i \bar{x}_{i,current}^2}} > \Delta$, return to step 2

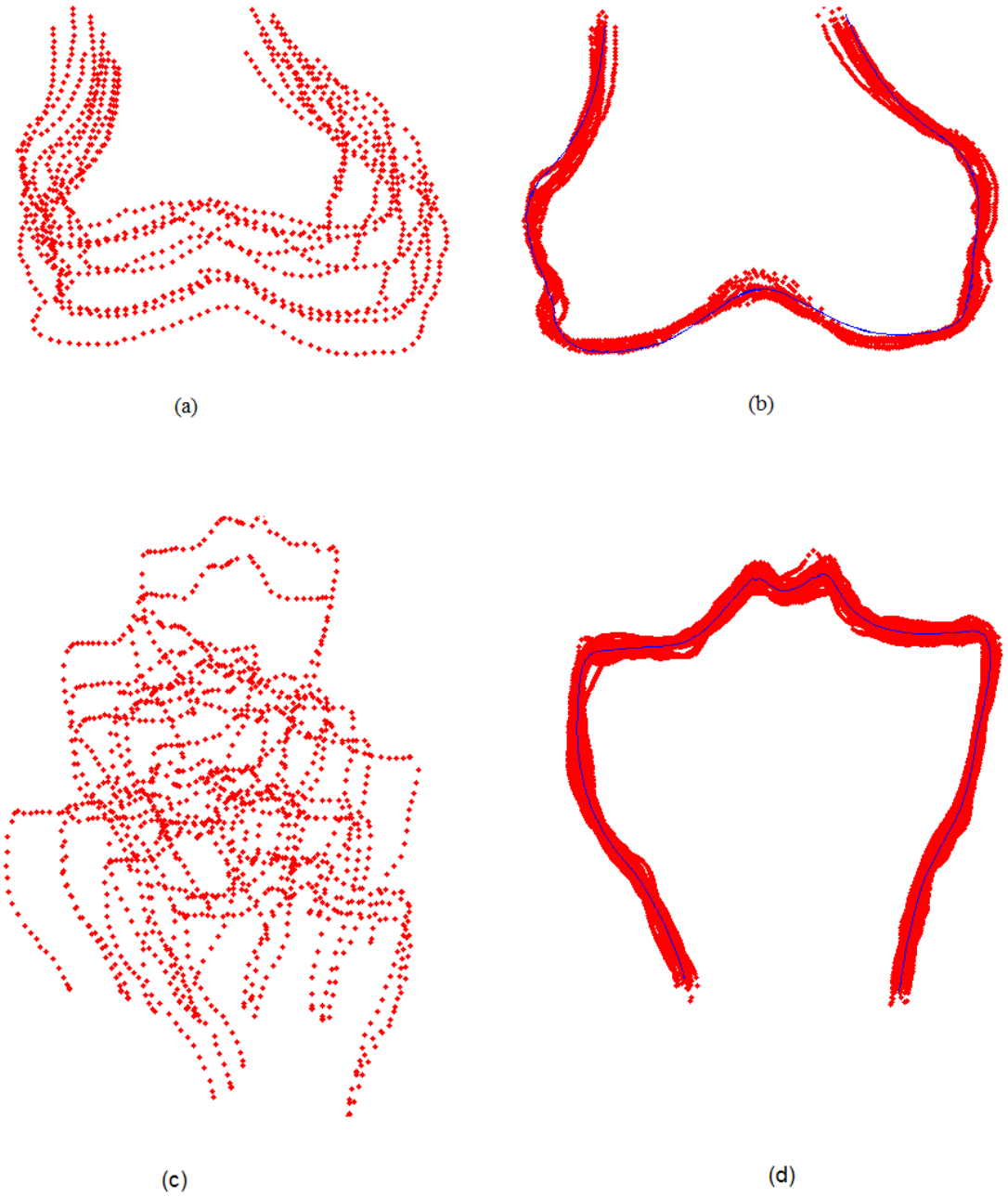


Figure 13. Training landmark points in (a), aligned in (b) with $\overline{X}$ shown in blue.

Figure 14 shows the cumulative variance as a function of the number of principal components for a typical model. The cumulative variance for the distal femur model approaches 99% at approximately 30 principal components.

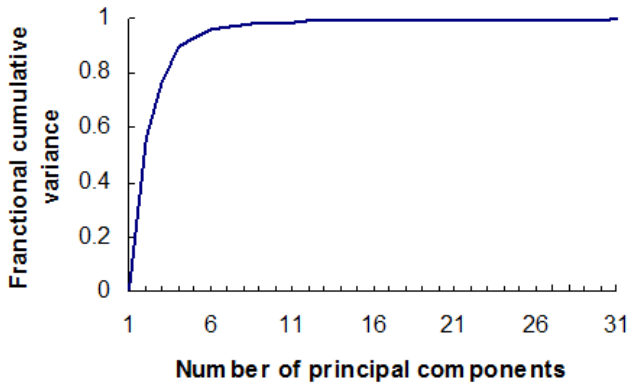


Figure 14. Average cumulative variance as a function of the number of principal components.

Figure 15 shows the square root of eigenvalues sorted by size (dotted line) together with components of one individual vector b_j. Since after the first few eigenvectors the variance becomes very small, the first *m* largest Eigen modes have been taken for building an active shape model that explains the variability of the contour shape.

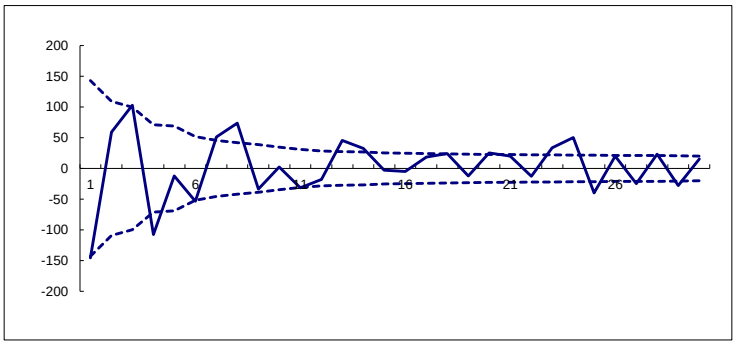


Figure 15. Statistics of shape deformation. The dotted line represents the square root of eigenvalues sorted by decreasing size. The continuous line corresponds to the components of an individual model coefficient, which describes the deviation of the shape from the mean shape.

B. Segmentation

Two commonly used segmentation strategies after spectral clustering are k-way segmentation [84] and hierarchical segmentation [86]. K-way segmentation is improved from recursive 2-way cut. It uses a clustering algorithm such as K-means to create a hard partition of the image. Unfortunately, this can lead to an incorrect segmentation as large uniform regions in which the eigenvectors vary smoothly are broken up. Hierarchical segmentation [86] treats each eigenvector as an image, from which a hierarchical segmentation strategy for going from contour signals to a hierarchical region tree calculates the nonoverlapped regions based on these contour signals by using OWT [94] and then constructs a hierarchy of regions in the UCM [86]. However, this method search the whole image which is time consuming and it cannot avoid infeasible results without using prior knowledge. Using the knowledge learnt from ASM to search only in the region of interest is both more computational efficient and robust to noises. Thus, we propose to start searching from the mean model to search for candidate points using spectral clustering and constrain them in the feasible shape with ASM as our segmentation strategy.

Based on the knowledge from statistical shape model, maximal gradient response on the normal profile of each point of the mean model is searched in the spectral clustered image. The calculated direction of normal profile is filtered by Savitzky-Golay filter [95] to get a smooth profile. Fig. 16 shows the normal profile and the gradient on the filtered normal profile of a sample image.

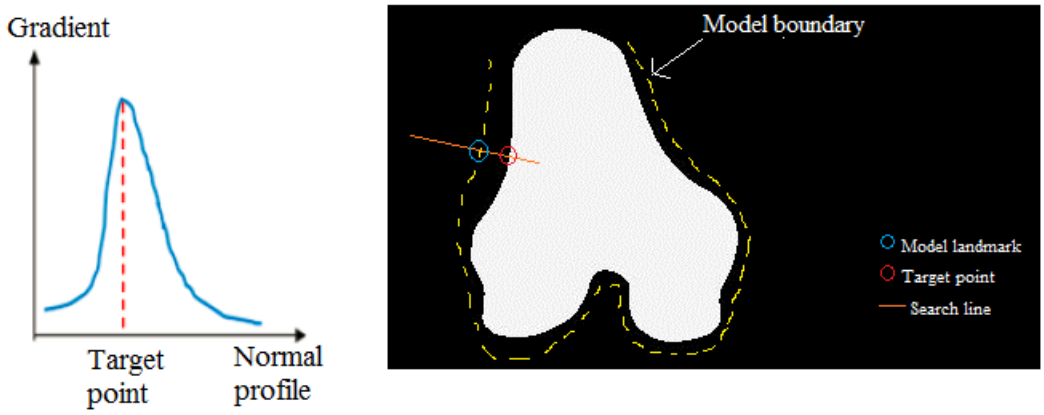


Figure 16. The gradient variation on the normal profile (search line) and the orientation of normal profile.

4.3 Experiments and Evaluation

In this section we designed experiments (1) to test how spectral clustering can enhance the image contrast; (2) to test how ASM can model the variation of bone structures; (3) to validate the proposed segmentation algorithm; (4) to test the robustness of the proposed algorithm.

Our data set comprises anteroposterior (AP) femur radiographs from 4,796 subjects suffering from osteoarthritis. All images were obtained from the public use datasets (version 0.E.1 clinical dataset) of the Osteoarthritis Initiative (OAI; online at www.oai.ucsf.edu)[25]. The images have been collected from different radiographic centers, resulting in large intensity differences due to the use of different X-ray and recording equipment. In addition, the presence of neighboring bones and soft tissue may result in mis-segmentation when these artifacts have stronger edge intensities in the images than the bone of interest. For these reasons, it is necessary to develop a robust automatic

segmentation system that can perform accurate segmentation for a wide variation of X-ray images.

These tests are based on a leave-one-out procedure. For each training set of a particular anatomical structure, we leave the testing bones out and compute a shape model based on the training examples. In the femur case, we used around 400 images for training and testing. We used 328 manually segmented contour as training set and the rest 80 femur contours and 20 tibia contours as testing set. The average error and deviation between the solution and the manual segmentation are computed.

The first experiment is designed to test the spectral clustering's ability to improve the contrast of the original x-ray image. Fig. 17 shows the first eight generalized eigenvectors from spectral clustering and compares them with the image enhanced by 3×3 Sobel operator. The intensity of both spectral clustering and Sobel operator images are normalized to [0,1]. The output from spectral clustering outperforms the conventional Sobel operator because it extracts only the most salient edges in the image while the latter fires at all the edges.

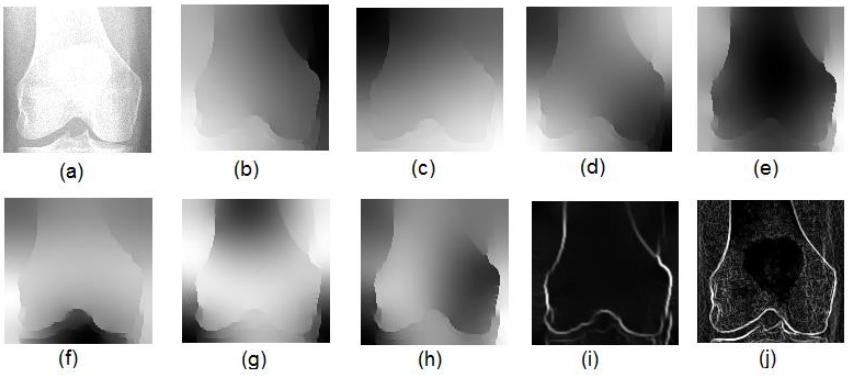


Figure 17. X-ray image enhanced by spectral clustering and Sobel operator. (a) original image; (b)-(h) first seven generalized eigenvectors of spectral clustering; (j) Maximum gradient response of the first seven eigenvector over orientations; (h) image enhanced by Sobel operator. The intensity of both spectral clustering and Sobel operator images are normalized to [0,1].

In the second experiment, shape parameter is limited within the range of $\left[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}\right]$ and the number of eigenvectors is limited to 30. Eigenvector is a set of displacement vectors, along which the mean shape is deformed. To stress this point, the first *m* eigenvector s has been plotted on the mean shape in Fig. 18, which result deformation of the mean shape. Since the limited number of eigenvectors and the limited value of shape parameters are used, we cannot expect the ASM based on the remaining examples to model the testing case perfectly. The average error is 0.2708 ± 0.1452 mm and the maximal error is 1.0996mm. These values represent lower limits on the errors obtained in the other experiments.

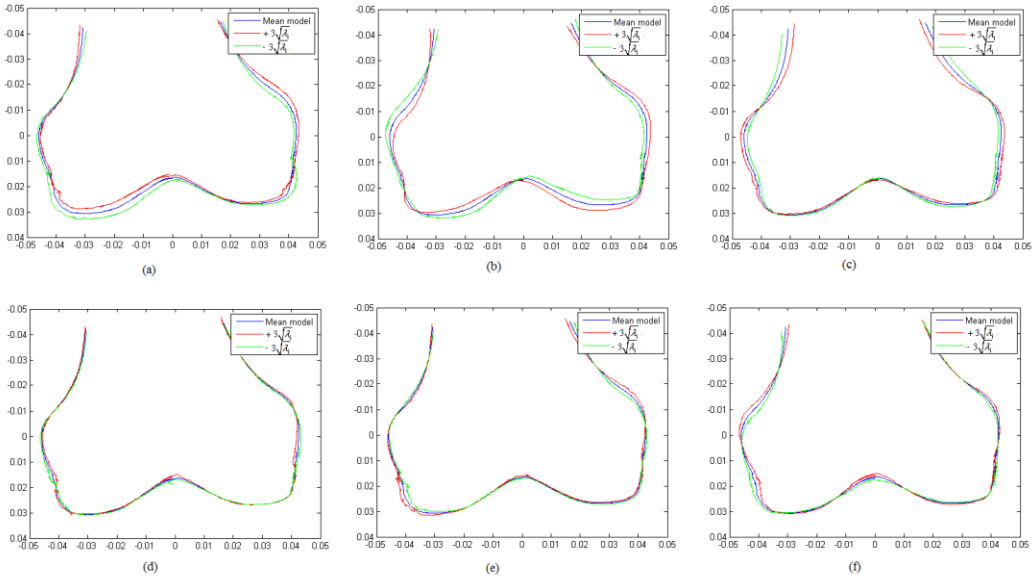
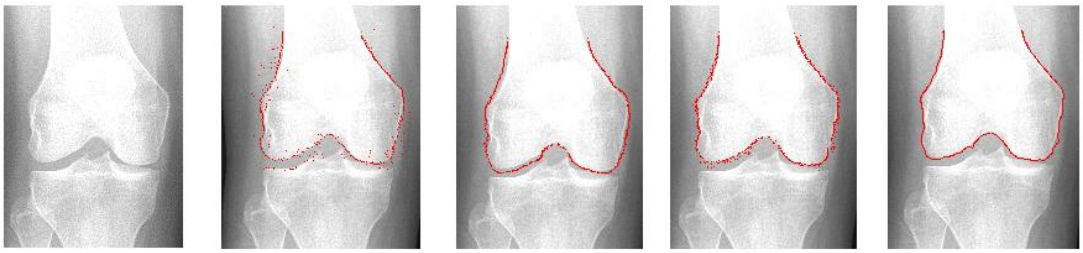


Figure 18. Mean shape deformation using 1st, 2nd… 6st principal mode; (a)-(f) is the first to sixth principal mode.

The next experiment tested the proposed algorithm with respect to its capability of finding the correct boundary of femur and tibia. Results for each step during the automatic segmentation are shown in Fig. 19. Fig. 19(a) is the first eigenvector in the spectral clustering, which shows the enhanced image with high contrast in the edges. Fig. 19(b) is the result after gradient search. Fig. 19(c) is the result after ASM constraint, which correct the outliers to feasible position. As we can see from the bottom part of the femur in Fig. 19(c), the mistakenly segmented points in tibia is corrected by the ASM constraint. Fig. 19(d) is the result after relaxation.



(a) (b) (c) (d) (e)

Figure 19. An example of segmentation results by spectral clustering based ASM; (a) original image; (b) gradient search result; (c) result after ASM constraint; (d) result after relaxation; (f) final result.

Qualitative results for the automatic segmentation are shown in Fig. 20. In these images, the ground truth (as determined by an expert) are shown in green, and the results from the automatic segmentation are shown in red. Dice measure is used to compare the similarity between segmented and ground truth contour [96]. Dice similarity measure of the segmented femur contour is 0.9696 ± 0.0099 mm with the worst case 0.9263; that of segmented tibia contour is 0.9834 ± 0.0095 mm with the worst case 0.9539.

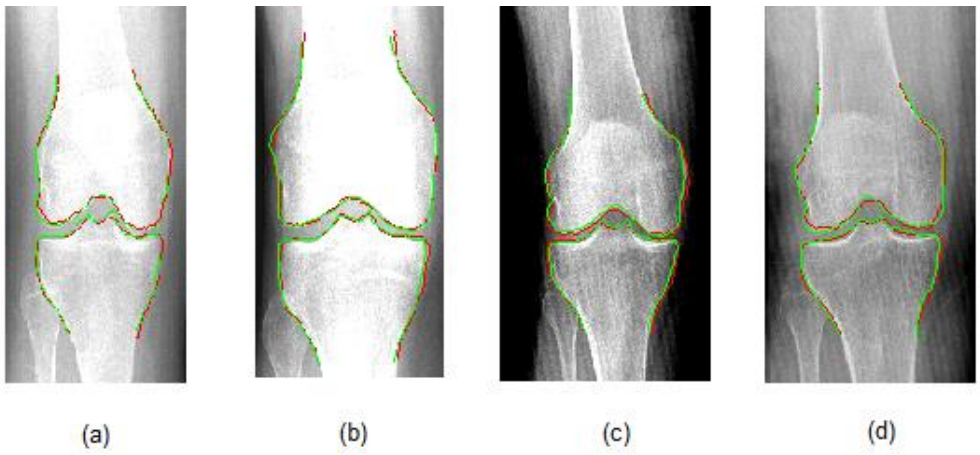


Figure 20. (a)-(d) show qualitative results with ground truth in green and results from automatic segmentation in red.

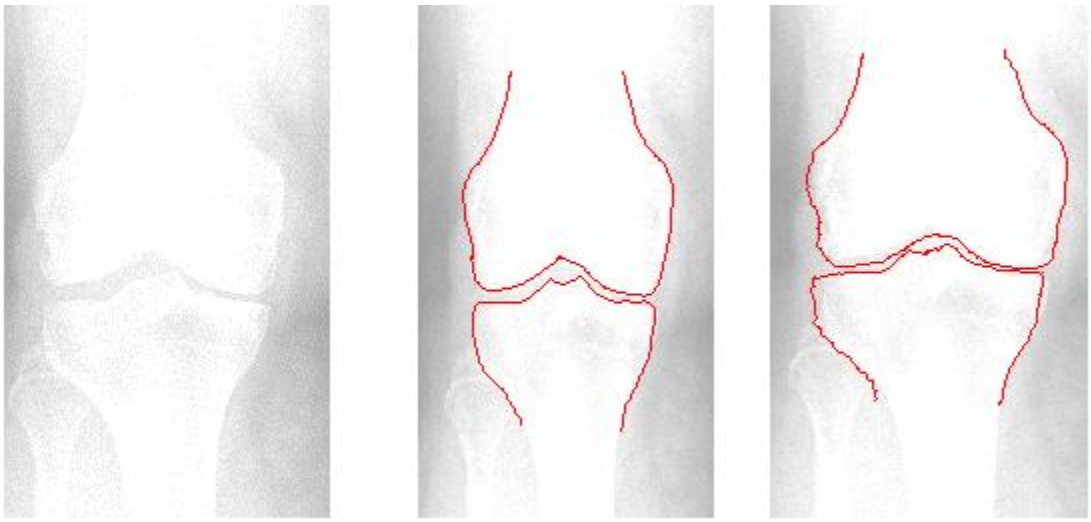
We compared the original ASM algorithms, the original ASM with relaxation, and the improved spectral clustering based search algorithm, with respect to their capability of finding the correct boundary of femur. This comparison is performed by computing the Root Mean Square (RMS) error between manual and automatic contours. Totally, eighty images were used for femur testing and twenty images for tibia testing. Ideally, these distances are zero for a perfect fit, assuming that the automatic segmentation is the same as the manual segmentation. Table 1 shows the RMS error, its deviation and maximal error. The last experiment examines the robustness of the search algorithm with respect to the presence of neighboring bones and low contrast. The final segmentations with and without prior shape knowledge in Figure 21(b) and Figure 21(c) show that the shape prior avoids infeasible points when miss segmentation occurs at the neighboring bone. Figure 22(a)-(c) shows low contrast X-ray images with low light strength and Figure 22 (d)-(f) shows images with high light strength. In both cases, spectral clustering based ASM algorithm compensate missing information despite of noise and low image contrast; while gradient

based method is very sensitive to noises.

Table 1. Compare the performance among original ASM, original ASM with relaxation, and spectral clustering based ASM.

	RMS error (mm)	Std (mm)	Maximal error (mm)
ASM, femur	0.6194	0.1727	1.1699
ASM with relaxation, femur	0.5964	0.2055	1.1271
Spectral clustering based ASM, femur	0.4766	0.1343	0.9035
Spectral clustering based ASM, tibia	0.5281	0.1816	1.1289

The last experiment examines the robustness of the search algorithm with respect to the presence of neighboring bones and low contrast. The final segmentations with and without prior shape knowledge in Fig. 21 (b) and 21(c) show that the shape prior avoids infeasible points when miss segmentation occurs at the neighboring bone.



(a) (b) (c)

Figure 21. Segmentation with shape prior on an image of femur occlusion from the neighboring bone: (a) original image, (b) segmentation with shape prior, (c) segmentation without shape prior.

4.4 Conclusions

This research presents a robust automatic segmentation algorithm combining statistical shape models and spectral clustering with the objective of accurately segmenting the distal femur and proximal tibia in X-ray images. The results demonstrate a significant improvement in addressing previously unanswered problems such as the existence of neighboring bones and occlusions due to implants, image artifacts or inconsistencies, significant contrast variations, and inter-patient anatomical femur/tibia shape variations. For applications of kinematics analysis, joint space width assessment, and preoperative planning, such an accurate robust segmentation method is required.

CHAPTER V
INITIALIZATION

X-ray video fluoroscopy along with 2D-3D registration techniques are widely used to study joints in-vivo kinematic behaviors. These techniques, however, are generally very sensitive to the initial alignment of the 3D model. In this section, an automatic initialization is presented for 2D-3D registration of medical images. Contour of the knee bone was first automatically extracted from 2D X-ray image. Shape descriptors were calculated by normalized elliptical Fourier descriptors (NEFD) to represent the contour shape. The optimal pose was then determined by a hybrid classifier combining k-nearest neighbors (KNN) and support vector machine (SVM). The proposed method can be easily utilized for 2D-3D image registration on various medical objects and imaging modalities. In the following part, Section 5.1 is the introduction, Section 5.2 describes the detailed approach, Section 5.3 shows the experiments and results.

5.1 Introduction

In vivo kinematic analysis using x-ray video fluoroscopy is applied in the evaluation of joint kinematics for both implanted and normal joints [4, 5, 77]. Fig. 22 shows fluoroscopic imaging in the knee during gait analysis. X-ray video fluoroscopy is a sequence of x-ray images, where the metallic implants and the bone appear much darker than the soft tissues surrounding them, allowing for direct observation and analysis of their movement under dynamic and weight bearing conditions. Moreover, fluoroscopy is noninvasive and relatively low risk to the patient. A typical one minute protocol gives the patient a radiation exposure on the order of 0.6 to 1.8 “rad equivalent man” (rem) [4].

However, two dimensional (2D) fluoroscopic images do not provide three dimensional (3D) information. 3D CAD/CT/MRI model is time-consuming (MRI) to generate and produces more radiation (CT) on patients, but it provides 3D insight and improves visualization of anatomical structures. Therefore, it is necessary to fuse the information of the 2D images and 3D volume data by 2D-3D registration.

Figure 22. Fluoroscopic imaging in the knee.

In clinical practice, it is desired to keep the number of X-ray fluoroscopic images to a minimum due to cost, acquisition and computation times and due to radiation exposure constraints. Single plane fluoroscopic imaging has been commonly used to study joint kinematics. Mahfouz et al. illustrated that single plane registration can be achieved with an in-plane accuracy of less than 0.091 mm in translation, with an out-of-plane error of 1.376 mm [5]. Different groups have investigated the use of dual fluoroscopy as a method for increasing the out-of-plane registration accuracy [97]. However limited studies have investigated the differences between the two methods using consistent 2D-3D matching

setup [98].

The key step in 2D-3D registration is to determine the relationship between the 2D X-ray image coordinate system and the 3D model coordinate system. This can be achieved by manual registration of 3D data to be fitted to the 2D x-ray image. This method, however is labor intensive and prone to both inter and intra user variability. An alternative is the use of paired point analytic registration with point data obtained from either skin or bone fiducial markers [99-101]. However skin markers are susceptible to error due to undesired motion between markers and the underlying bone, and bone markers are invasive which limits their use. A third method of 2D-3D image registration is to iteratively adapt the unknown pose as to maximize a similarity measure which reflects the quality of the registration. A large body of work of 2D-3D registration based on 3D model registration to radiographs [5, 6, 16, 35] and 3D kinematic analysis from fluoroscopic images [102] have shown satisfactory registration accuracy. Existing 2D-3D registration methods are often limited by a small capture range[103]. Due to the non-convex nature of the similarity measure, such methods can suffer from inaccurate registration results when the initial pose of the 3D model is outside the capture range. The objective of this paper is to develop a method that provides a large capture range for the 2D-3D registration process by comparing a query shape in 2D image with a pre-computed training model to determine the initial pose of the 3D model. The main advantage of the proposed method is the ability to search in a large range of possible solutions making the methods independent of initialization. Moreover, since the pose estimation processes using X-ray fluoroscopy in various planes are independent from each other, our method can be applied to both monoscopic and stereoscopic registration.

There is a wide variety of published literature on the topic of pose estimation based on template matching [75, 104-106]. Khotanad and Liou [105] developed a neural network based system for recognition and pose estimation of an unoccluded three-dimensional object from any single two-dimensional perspective view. The neural network (NN) classifier used in this paper may suffer from multiple local minima and is more prone to over fitting compared with support vector machine (SVM). Unlike NNs, SVMs have a simple geometric interpretation and give a sparse solution. Moreover, the computational complexity of SVMs does not depend on the dimensionality of the input space, which is not the case for NN. Therefore, SVMs often outperform NN in practice where SVMs are less prone to over fitting and yield global and unique solution [107].

Banks and Hodge [75] used the template matching method to measure the pose of knee prostheses by matching the projected silhouette contour of the prosthesis against a library of shapes representing the contour of the object over a range of possible orientations. Hoff [106] matched the silhouette of the prostheses components against a library of images to estimate the position and orientation of the component. Both methods directly matched the projected 3D model image with a template library of implant models. Moreover, they applied the pose estimation to implant models, which has limited shape variation when compared to bone anatomy.

Direct implementation of the template matching method would require an exhaustive computation time and usage of memory by comparing the distance between the query and the entire training examples. Instead, we used a hybrid classifier to determine the pose of the 3D model so as to avoid a blanket search. More specifically, the hybrid classifier combines k-nearest neighbors (k-NN) and support vector machine (SVM), where

k-NN serves as an attention mechanism, generating an area of interest for the further local search by SVM [108, 109]. A set of SVMs, each trained to a smaller range of poses, decomposes the pose estimation problem across a large range of poses into a set of sub-problems. Thus, the proposed initialization method is more time and space efficient than the conventional direct template matching methods. Moreover, our method can be easily adapted for various anatomical or implant structures and various imaging modalities.

The remainder of the paper is organized as follows. The approaches are described in Section 2. Section 3 provides the experiments and their results. Finally, Section 4 concludes the paper and gives future prospects.

5.2 Approach

5.2.1 Problem Statement

The goal of the proposed method is to estimate the poses of the 3D model from single plane fluoroscopic x-ray image based on a precomputed training set of 2D contours. This method helps kinematics in that it can determine 3D in-vivo, weight bearing kinematics of the knee joint from single plane fluoroscopic image. It also helps biomechanics in that it determines the orientation of the femur with respect to the tibia so as to determine the forces and torques acting on the knee joints.

Let $S \in \Re^3$ be the smooth surface of the 3D model, and denote $X = [x, y, z], X \in S$ to be the corresponding spatial coordinate. Let $X_0 \in \Re^3$ and $S_0 \in \Re^3$ be the coordinates and surface in the 3D world, respectively. We can locate the S in the camera reference frame via the transformation T such that $S = T(S_0)$ and the corresponding pointwise

expression

$$X = T(X_0) = RX_0 + t \tag{5-1}$$

where $R \in SO(3)$ is a rotation matrix $R = R_\gamma * R_\beta * R_\alpha$, $R_\alpha = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}$,

$R_\beta = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}$, $R_\gamma = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $t \in \Re^3$ is a translation vector,

$t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$. X_0 is the coordinates of the 3D model in the 3D world.

The proposed method consists of three steps: (1) Shape extraction to extract the 2D contour of the object of interest in an X-ray fluoroscopy. (2) Shape representation to represent the shape by normalized elliptical Fourier descriptors (NEFD). (3) Classification by a hybrid classifier to estimate the pose of the 3D model. The system framework is illustrated in Fig. 23.

5.2.2 Shape Extraction

The contour of the object of interest in the X-ray fluoroscopy is extracted by an active shape model (ASM) based segmentation algorithm [3]. ASM represents the global shape constraints with the dominant shape variation in the training set. The benefit of ASM

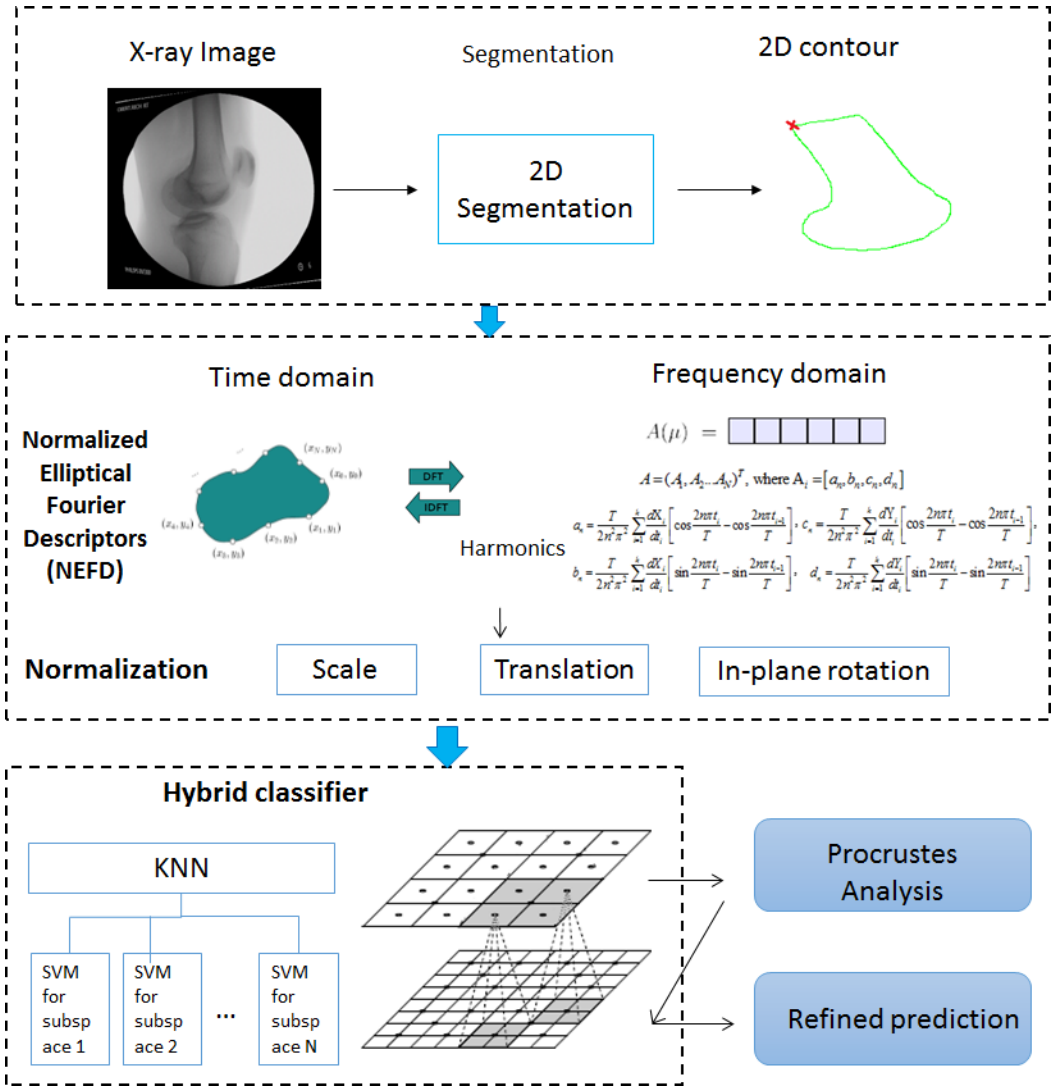


Figure 23. Flow chart of the initialization system.

is that it constrains the search in feasible shape so that mis-segmentations such as skin segmentation or segmentation of neighboring bones and foreign occlusions can be avoided. The training data are the contours extracted from the binary images, which are the 3D surface mesh models projection onto the 2D plane at different poses, yielding 1,887 training contours. The contours are then aligned by finding corresponding points of each contour using statistical atlases, where training contours with known point distribution are deformed to match one another [110-112]. To segment the X-ray image, we start from the average contour in the training data set; search on the normal profile of each contour point for the greatest gradient; use ASM to constraint the contour in feasible shape; and then relax the constrained points for the segmented contour. Relaxation is done by a gradient search in the normal profile once again with a half profile length. The contour is closed by drawing a line perpendicular to the major axis of the bone, the direction of which can be represented by the first principal component of the PCA of the contour points. The flowchart of the segmentation algorithm is illustrated in Figure 24. An example of segmentation results are shown in Fig. 25.

5.2.3 Shape Representation

The shape representation is based on normalized elliptical Fourier descriptors (NEFD) developed by Kuhl [113]. Contours are first normalized by in-plane rotation, translation, and scale. NEFD represents the object shape in a very compact manner and is therefore time and space efficient.

A. Calculate shape descriptors of the contour

The main idea of the elliptical Fourier analysis is to approximate a closed contour

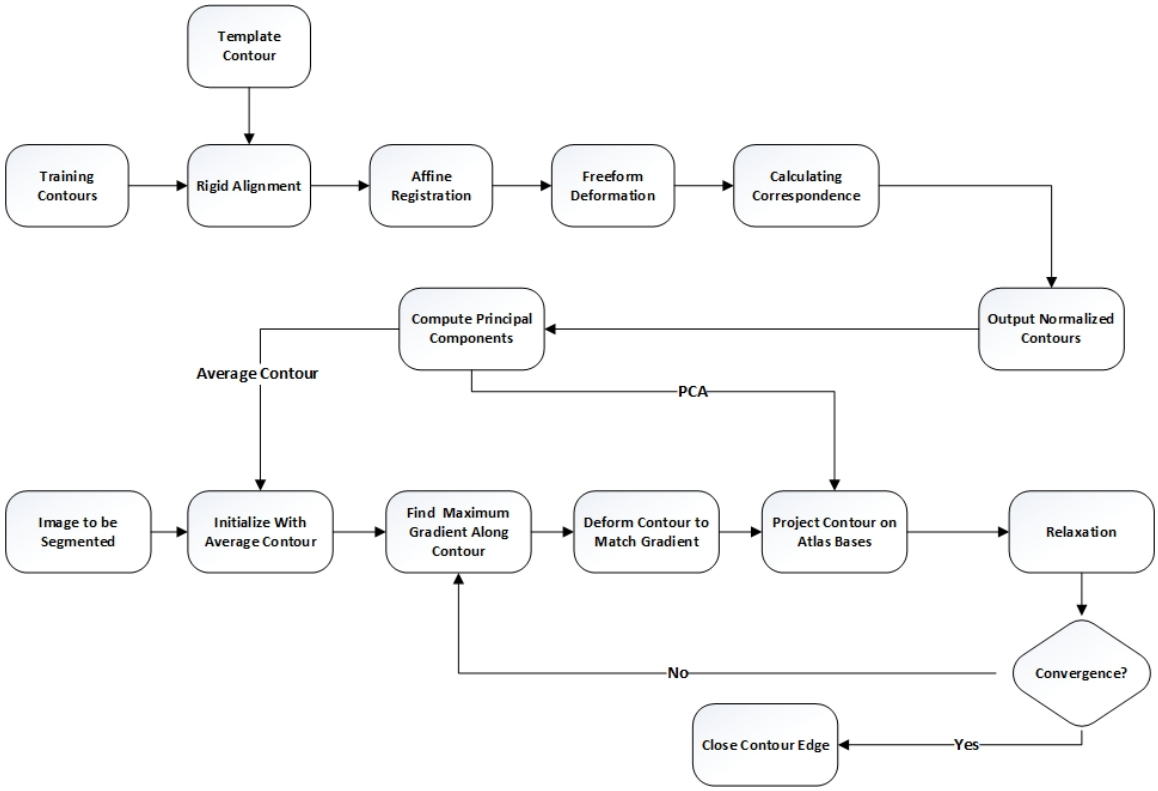


Figure 24. Flow chart of 2D segmentation algorithm.

Figure 25. Segmentation result of femur and tibia in an X-ray image.

as a sum of elliptic harmonics. N harmonics are used to identify the closed contour of k elements. Four coefficients are used for each harmonic, as given by

$$a_n = \frac{T}{2n^2\pi^2}\sum_{i=1}^{k}\frac{dX_i}{dt_i}\left[\cos\frac{2n\pi t_i}{T} - \cos\frac{2n\pi t_{i-1}}{T}\right], \tag{5-2}$$

$$b_n = \frac{T}{2n^2\pi^2}\sum_{i=1}^{k}\frac{dX_i}{dt_i}\left[\sin\frac{2n\pi t_i}{T} - \sin\frac{2n\pi t_{i-1}}{T}\right], \tag{5-3}$$

$$c_n = \frac{T}{2n^2\pi^2}\sum_{i=1}^{k}\frac{dY_i}{dt_i}\left[\cos\frac{2n\pi t_i}{T} - \cos\frac{2n\pi t_{i-1}}{T}\right], \tag{5-4}$$

$$d_n = \frac{T}{2n^2\pi^2}\sum_{i=1}^{k}\frac{dY_i}{dt_i}\left[\sin\frac{2n\pi t_i}{T} - \sin\frac{2n\pi t_{i-1}}{T}\right] \tag{5-5}$$

where X_i and Y_i are X and Y coordinate values; a_i and b_i represents the projection on the semi major and semi minor x-axis, similarly c_i and d_i represents the projection on the semi major and semi minor y-axis. $T = \sum_{i=1}^{m} dt_i$ is the length of the Freeman chain [18] of the closed contour and t_i is the length of i elements of the chain, and m is the number of elements in the chain.

The inverse process allows identification of the closed contour from the N coefficients, as given by

$$X_i = X_c + \sum_{n=1}^{N} a_n \cos\frac{2n\pi t_i}{T} + b_n \sin\frac{2n\pi t_i}{T} \tag{5-6}$$

$$Y_i = Y_c + \sum_{n=1}^{N} c_n \cos\frac{2n\pi t_i}{T} + d_n \sin\frac{2n\pi t_i}{T} \tag{5-7}$$

where X_c and Y_c represent the coordinates of the centroid.

The most significant features of the contour are captured in the lower frequency

terms of the Fourier descriptors. Therefore, a more concise representation of the contour can be generated by eliminating the high-frequency terms. As can be seen in Figure 26 (a) where root mean square (RMS) error between the reconstructed and the original contour of femur drops as the number of harmonics increases. Since the reconstruction error remains steadily low when the number of harmonics reaches forty, forty harmonics were found to provide accurate representation for the training shapes, as seen in Figure 26 (a). The comparison between the original and reconstructed contours from Fourier descriptors are shown in Figure 26 (b).

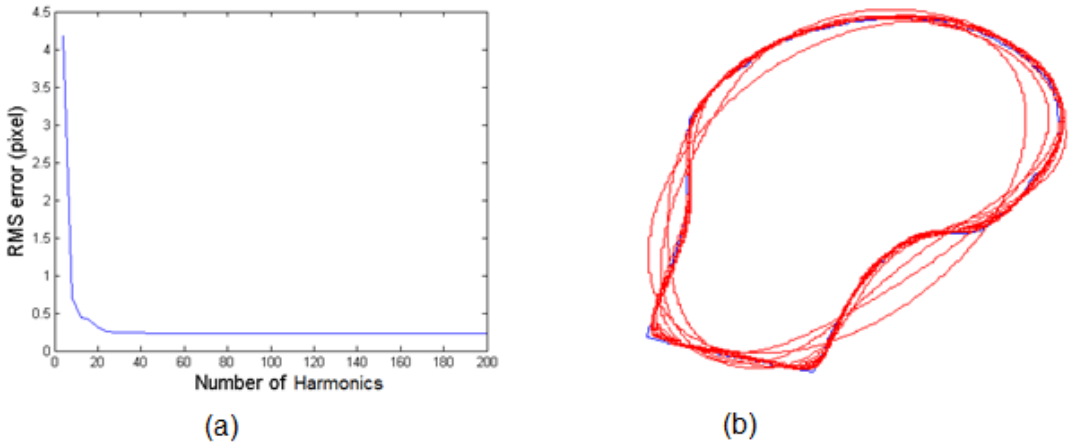


Figure 26. Reconstruction error of NEFD (a) RMS error of reconstructed femur contours with respect to the number of harmonics. (b) Reconstructed femur contours are in red with the number of harmonics from 1 to 50. The original contour is in blue.

B. Normalization

The centroid of the contour is moved to the origin to normalize translation, calculated by

$$X'_i = X_i + \frac{1}{k}\sum_{n=1}^{k} X_n \tag{5-8}$$

$$Y'_i = Y_i + \frac{1}{k}\sum_{n=1}^{k} Y_n \tag{5-9}$$

Because the starting point can be chosen randomly and such a choice impacts the coefficient values, it is necessary to have a representation which is invariant to the starting point on the contour. This is achieved by rotating the elliptic coefficients until a shift phase is equal to zero, as given by

$$\begin{bmatrix} a_n^* & b_n^* \\ c_n^* & d_n^* \end{bmatrix} = \begin{bmatrix} a_n & b_n \\ c_n & d_n \end{bmatrix} \begin{bmatrix} \cos n\theta_1 & -\sin n\theta_1 \\ \sin n\theta_1 & \cos n\theta_1 \end{bmatrix} \tag{5-10}$$

where $\theta_1 = \frac{1}{2} arctg \frac{2(a_1 b_1 + c_1 d_1)}{a_1^2 + c_1^2 - b_1^2 - d_1^2}$ is the phase's shift from the first major axis.

The rotational invariant can be achieved by aligning the major axis of the first harmonic of each image to the x-axis. The coefficient is normalized by the following equation:

$$\begin{bmatrix} a_n^{**} & b_n^{**} \\ c_n^{**} & d_n^{**} \end{bmatrix} = \begin{bmatrix} a_n^* & b_n^* \\ c_n^* & d_n^* \end{bmatrix} \begin{bmatrix} \cos\phi_1 & \sin\phi_1 \\ -\sin\phi_1 & \cos\phi_1 \end{bmatrix} \tag{5-11}$$

where $\phi_1 = arctg \frac{c_1^*}{a_1^*}$

The size invariant is achieved by dividing the length of the semi-major axis, *L*, of the first harmonic:

$$L = \sqrt{a_1^{*2} + c_1^{*2}} \tag{5-12}$$

The invariance of translation, starting point, rotation, and size allows an efficient representation of all the possible shapes.

5.2.4 Estimate Out-of-plane Rotation by a Hybrid Classifier

A hybrid classifier was used to estimate the out-of-plane rotation. We consider the estimation of the out-of-plane rotation in the framework of measuring similarities or equivalently distances to the training shapes by a combination of k-nearest neighbor (KNN) and support vector machine (SVM) classifiers. KNN is a non-parametric classifier. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors [114]. While KNN is natural in this kind of problem, it suffers from high variance in the case of limited sampling. Moreover, since nearest neighbor classifier uses all the instances in the training set, it is computationally expensive. The alternative is SVM, which models the nonlinear structure of the data by mapping the data onto the high dimensional kernel space and only seeks to model the boundary of classes. But SVM involves time-consuming optimization and computation of pairwise distances, which makes it difficult to classify a large amount of classes. We thus combine KNN and SVM to improve the overall performance.

Since K-NN is used for the coarse search, only 5° increment instances are used. Euclidean distance (Ed) is used to measure similarity between objects and is computed as given below,

$$Ed(A,B) = \sqrt{\sum_{i=1}^{N-1}(s_{Ai} - s_{Bi})^2} \tag{5-13}$$

where $Ed(A,B)$ is the Euclidean distance between instance descriptors A and B, s_{Ai} and s_{Bi} are the i^{th} NEFD of instance A and B, respectively.

SVM provides a good generalization for pattern classification problems without incorporating problem domain knowledge. Recall that in a two-class SVM classification

problem, we want to find a hyper plane that separates the two classes with maximal margin [115]. When it is not possible to linearly separate the two classes, kernel function is used to map the data into high dimensional space where the data is linearly separable. Given training vectors $x_i \in R^n$, i =1,...,l, and their label vector $y \in R^l$ such that $y_i \in \{1,-1\}$ and the mapping $\phi(x_i)$, SVM solves the following optimization problem.

$$\begin{aligned} &\min_{\omega,b,\xi} \quad \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{l}\xi_i \\ &s.t. \quad \begin{cases} y_i\left(\omega^T\phi(x_i)+\omega_0\right) \geq 1-\xi_i, \\ \xi_i \geq 0, i=1,...,l, \end{cases} \end{aligned} \tag{5-14}$$

where ω is a weight vector and ω_0 is the threshold, $\phi(x_i)$ maps x_i into a high-dimensional space and C>0 is cost coefficient, which represents a balance between the model complexity and the approximation error. When the constraint conditions are infeasible, slack variables ξ_i can be introduced.

There are different choices of kernels depending on their application at hand. An appropriate selection of the kernel may drastically affect the final classification performance. The RBF kernel (or Gaussian kernel) is the most commonly used kernel. Glasner reported a good performance of RBF kernel [116] in the pose estimation problem. We also compared the classification accuracy on the same set of testing data between different kernels as shown in Table 2.

Table 2: Classification accuracy of different kernel functions. The success rate is calculated with the error threshold set as 4 degree

KERNEL TYPE	Success rate
Linear	65.23%
Sigmoid	65.14%
Polynomial	64.52%
RBF	95.98%

We used Gaussian Radial-Basis Function (RBF) kernel, $\Phi(x_i \cdot x_j) = \exp(-\gamma \left\| x_i - x_j \right\|^2)$.

Penalty parameter C in (14) and γ in kernel function are two of the most important parameters. C, a regularization parameter, affects the tradeoff between maximizing the margin and minimizing the training error. Both C and γ determine the number of support vector. Since SVM only models the boundary between each class, the number of support vectors impacts the training time and accuracy. γ also affects the amplitude of the Gaussian function and, therefore, affects the generalization ability of SVM [117]. The technique used to determine its optimal parameters is a grid search using a cross-validation [107, 118] that searches the optimal parameters by a coarse grid search with an exponentially growing sequence of (C, γ) with $C=2^{-5}, \ldots, 2^{15}$ and $\gamma=2^{-5}, \ldots, 2^{15}$ followed by a finer grid search to find the one giving the highest accuracy.

Since SVM is a two-class classifier, commonly used methods to extend the two-class classifier into multi-classes include one-against-all method, one-against-one [119], and directed acyclic graph [120]. We use the one-against-one method [119] in this paper. Scaling of the feature vectors is also one of the important factors affecting performance [107]. Training set was normalized by scaling it into [0, 1] and then test data was normalized with the same degree of scaling.

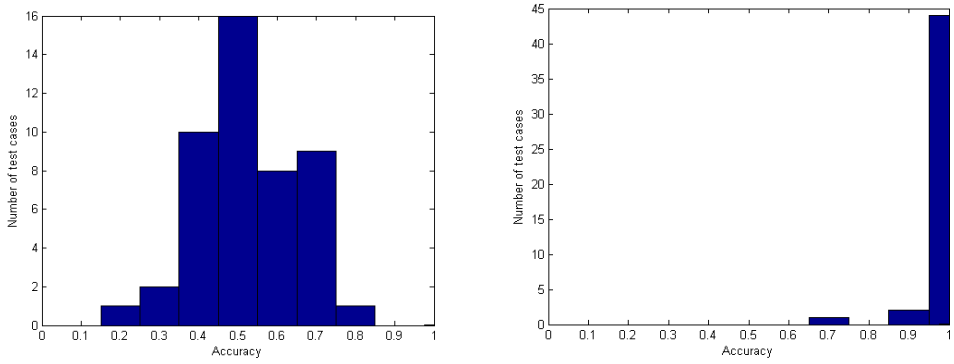
5.2.5 Estimate In-plane Rotation

(5-11) in Section 5.2.3 provides an approximation of the in-plane rotation by aligning the major axis of the first harmonic of each contour, which is a rough approximation of the real shape. Therefore, a more accurate method is discussed here to

determine the in-plane rotation by the angular difference between the best matching contour in the training set and the testing contour through Procrustes Analysis [121].

$$\theta_{z_est} = \theta_{z_input} - \theta_{\Delta} \tag{5-15}$$

Before Procrustes analysis, corresponding points across contours must be established using statistical atlases, same as the corresponding point procedures discussed in Section 5.2.2. The result of in-plane rotation estimation is shown in Fig. 27.



(a) (b)

Figure 27. Compare the results of out-of-plane rotation by KNN and in-plane rotation by Procrustes Analysis. (a) Histogram of estimation of x and y axis rotation by KNN. (b) Histogram of estimation of z axis rotation by Procrustes analysis.

5.2.6 Pose Estimate Algorithm

Observe from the results of in-plane rotation estimation in Fig. 27 that even when the out-of-plane rotation estimation has large error, the following estimation of in-plane rotation successes in 90% of the cases.

Therefore, a two-stage pose estimation algorithm is designed. During training, a set of SVMs is built with each SVM is trained on the training data with fixed in-plane rotation (z axis). A diagram of the set of SVMs is shown in Fig. 28. During training, KNN is first

used to obtain a coarse estimation followed by an in-plane rotation estimation by Procrustes Analysis. This stage is used to select the correct SVM for the next stage. In the second stage, a selected SVM is used for out-of-plane rotation estimation, and followed by an in-plane rotation estimation by Procrustes Analysis. The detailed steps are shown in Algorithm 2.

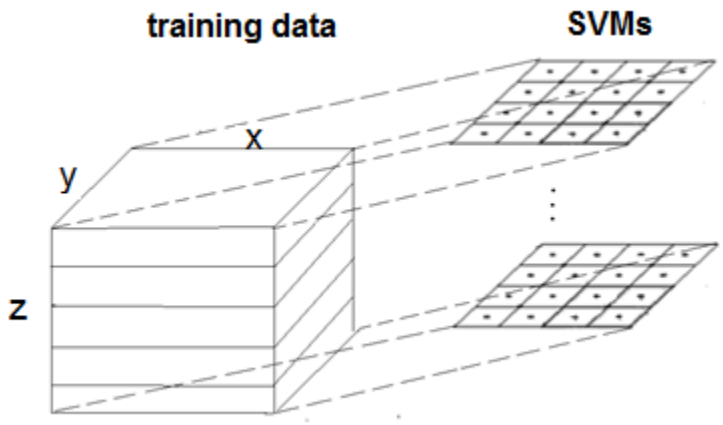


Figure 28. A diagram of SVMs set

Algorithm 2 Pose estimation algorithm

Input: Fourier Descriptors of the query 3D model (a_1,b_1, c_1,d_1,……a_n, b_n, c_n, d_n)
Output: Estimated pose: x_r, y_r, z_r
Training
1: for i = z_{r0}:z_{rt}
2: Train SVM on the poses $x_r \in [x_{r0}, x_{rt}], y_r \in [y_{r0}, y_{rt}], z_r = i$;
3: end
Testing
4: Use KNN only to predict the x and y axis pose (xr1, yr1);
5: Use Procrustes analysis to predict the z axis pose (zr1);
6: Predict the new x and y axis pose (xr2, yr2) by svm, which is selected as zr= zr1
7: Predict the new z axis pose (zr2) by Procrustes analysis

5.3 Experiments and Results

Three sets of experiments were performed to validate the proposed method. In the first set of experiments (Section 5.3.2) computer simulated images were used to estimate the 3DOF transformation parameters (rotation in x, y and z axis). A comparison was made among the accuracy of three different classifiers. In the second set of experiments (Section 5.3.3), we further tested the initialization method with x-ray images. In the third set of experiments (Section 5.3.4), the initialization of 6DOF transformation parameters (translation in x, y and z axis; rotation in x, y and z axis) was followed by a feature-based 2D-3D registration algorithm. A comparison of performance between initialized and uninitialized registration was made to test whether the proposed method leads to improved registration.

5.3.1 Training Data

William M. Bass Donated Skeletal Collection of 100 femurs and tibias from the University of Tennessee's Anthropology Department was used. Computed tomography (CT) scans of the knees were made at levels ranging from 120 mm proximal to the joint to 120 mm distal to the joint. These scans were made at 1-2 mm intervals and the volumetric data of the knee joint was constructed at 0.5 mm interpolation in the transverse plane. Segmentation of the CT-scanned bone was automatically performed by applying a thresholding filter to the slices which isolated the bone from the background. Manual intervention was conducted only when the thresholding filter failed. An example of the 3D models is depicted in Fig. 29.

Figure 29. 3D models of knee bone

To generate the training set, the 3D models were rendered on the 2D plane by a software developed by the authors [5] using a 3D graphic library (Open Inventor, TGS, San Diego, CA). A coordinate system was established with x axis pointing to the left, y axis pointing upwards, and z axis pointing out of screen. The impact of the perspective effect on the projected shape was analyzed by comparing the shape variation between two poses with a different translation and fixed rotation during deep knee bending (DKB). The two translational values were set by the first and last pose of the DKB respectively. As can be seen from Fig. 30 (a) and (b), neither the shape nor the NEFD varies significantly for these two poses. In addition, in-plane rotation does not lead to any variation in the projected shapes. Therefore, we represent the shape by out-of-plane rotation (around x and y axes) only; while the translation and in-plane rotation (around z axis) are normalized by NEFD.

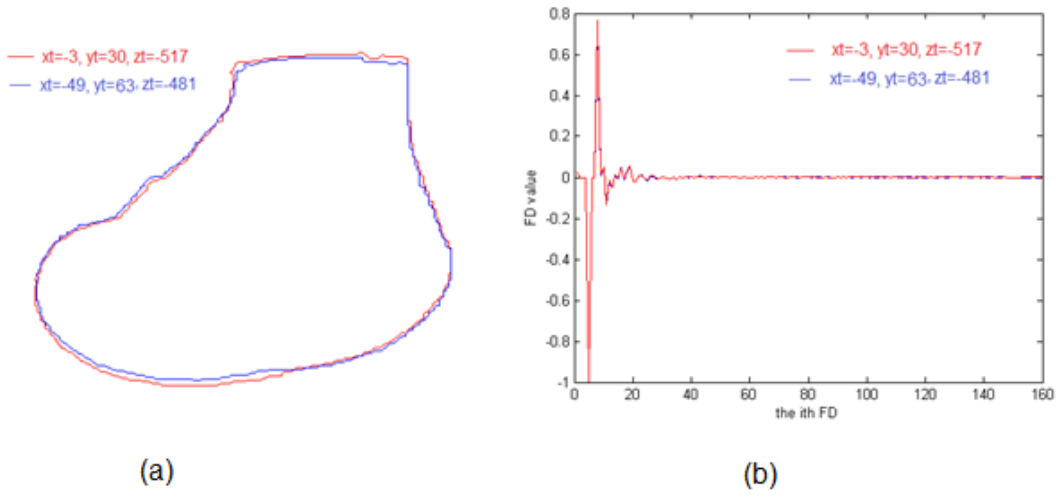


Figure 30. Perspective effect by comparing two poses with the greatest variation in translation and fixed rotation during deep knee bending. (a) Projected shapes; (b) NEFD of the projected shapes.

The 3D models were rotated at 1° increment about the x- and y-axis while fixing position and in-plane rotation. The rotational range is $x \in [-80, -40]$, $y \in [-175, -150]$ for the femur; and $x \in [80, 110]$, $y \in [-30, 0]$ for the tibia. At each orientation, the 3D models were rendered as a binary image so that only the silhouettes are visible.

5.3.2 Experiments on Simulated Data

Computer simulated data of femur, and tibia were generated by rendering the 3D models onto the 2D plane at known poses to validate the accuracy of the estimates provided by the hybrid classifier. The rotational parameters of the testing data were in the same range as training data in x and y axis, and varying in the range of 40° difference along the z axis. The increment was 1° for each of the three axial rotation. As an error measurement, we measure the absolute angular difference between the estimated pose and the ground truth data in each axis. Reported initial mis-registrations leading to successful registrations are

in the order of 4-11 mm mTRE [103, 122]. Since a rotation of 1° around one of the three axes will lead to an mTRE of 0.8-1.0 mm [123], estimates with an angular error smaller than 4° can be safely considered successful. The capture range was defined as the 95% success range as was proposed in [103].

Validation were performed to test the performance of the initialization method. For femur and tibia, ninety patients' data were used for training and ten patients' data for testing. Each patient's data contains an equal number of poses as class labels. Table 3 presents the estimates by the initialization method for femur and tibia. By setting the success threshold as 4 degree angular difference between estimated and gold standard poses [103, 122], 96% of the femur and 93% of the tibia estimation were successful.

Table 3: RMS error of synthetic data experiments with KNN-SVM classifier (90% patients for training and 10% patients for testing)

	Xr (deg)	Yr (deg)	Zr (deg)	Success rate
Femur	0.54 ± 0.92	1.57 ± 1.86	0.83 ±1.34	96%
Tibia	-0.14 ±1.34	-0.77 ±1.73	0.36 ±1.63	93%

5.3.3 Experiments on X-ray Images

Experiments were performed using x-ray fluoroscopic images acquired using a high-frequency pulsated video fluoroscopy unit. As the largest motion of the knee occurs in the flexion and anteroposterior translation, fluoroscopic imaging was conducted in the sagittal plane. The modern fluoroscopic systems allow video capture at a rate as high as 60 Hz. The fluoroscope is modeled as a perspective projection image formation model, which treats the fluoroscope sensor as consisting of an X-ray point source and a planar phosphor screen upon which the image is formed. The x-ray images had a resolution of 640 ×480 pixels. Screenshots of the x-ray images are shown in Fig. 31.

Figure 31. Fluoroscopic X-ray images.

Contours were automatically extracted using the segmentation algorithm mentioned in Section 5.2.2. The overall performance of the method was evaluated by calculating the absolute angular difference between the pose estimated using the proposed method and the ground truth, which is manual fitting. Twenty x-ray images in a series of fluoroscopic images during deep knee bending were used for validation from both datasets. The average error of femur, and tibia is shown in Table 4, where all the prediction poses are within the tolerance angle of four degrees. Overlaid of the estimated 3D model on the 2D X-ray image are shown in Fig. 32.

Table 4: Average Error of Experiments on x-ray images. Twenty x-ray images for each case

	Xr(◦)	Yr(◦)	Zr(◦)
Femur	0.60 ± 1.00	1.3 ± 1.49	0.80 ± 0.63
Tibia	1.50 ± 0.52	1.90 ± 0.32	1 ± 1.25

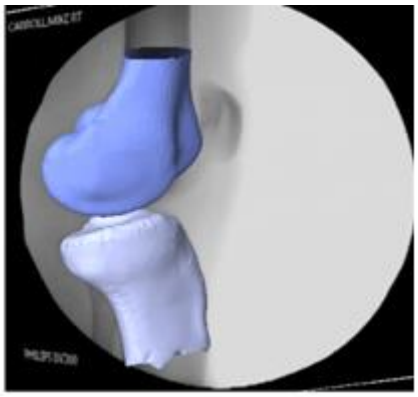


Figure 32. Overlaid of the estimated 3D model on the 2D X-ray image.

5.3.4 Feature-based 2D-3D Registration Experiments

Experiments were conducted to show the benefit of initialization for 2D-3D registration by comparing the results of initialized and uninitialized registration. The feature-based 2D-3D registration algorithm was implemented with a normalized cross-correlation similarity measure [5] and optimization was performed by Simulated Annealing. Registrations with a final error smaller than 1 degree were regarded as successes. The uninitialized registration used both the randomly selected starting pose and the middle of the angle range as its initial pose, while the initialized registration used the pose estimated by the proposed initialization method. Translation in x and y axis was estimated by the center of the contour. Translation in z axis was estimated by comparing the scale ratio of the testing contour with the one of known pose [32].

A comparison between the result of uninitialized and initialized registration was made using ten cases of femur and tibia, as shown in Table 5. Initialized registration had a success rate of 100%. In contrast the probability of success for uninitialized registration

remained as low as around 50% for the initial pose at the middle of the angle range and around 10% for a randomly selected initial pose.

Table 5: Results of Initialized and Uninitialized registration

Uninitialized registration(starting point at the middle of the angle range)					Initialization registration	
		Average error (◦)	Average error (◦) (obvious failure removed)	Correct registration %	Average error (◦)	Correct registration %
Femur	Xr	8.29 ±16.97	0.47 ±0.57	60%	0.26 ±0.98	100%
	Yr	1.23 ±2.33	0.25 ±0.4		0.09 ±0.18	
	Zr	2.08 ±3.52	0.63 ±0.86		0.46 ±0.96	
Tibia	Xr	0.82 ±1.18	0.26 ±0.45	50%	0.01 ±0.17	100%
	Yr	0.36 ±0.69	0.09 ±0.16		0.01 ±0.03	
	Zr	4.34 ±7.62	0.85 ±1.31		0.27 ±0.17	
Uninitialized registration (random starting point)						
Femur	Xr	8.66 ±14.41	1.43 ±2.23	20%		
	Yr	9.46 ±13.59	0.00 ±0.00			
	Zr	26.33 ±37.36	1.46 ±2.48			
Tibia	Xr	20.71 ±20.87	0.8 ±0.7	0%		
	Yr	7.96 ±11.37	0.9 ±1.13			
	Zr	12.85 ±17.36	1.4 ±1.56			

5.3.5 Speed of the System

The time required for pose estimation was about five seconds on average and the time for training SVM models was about 15 hours. The method was implemented using a combination of C++ and MATLAB and a 2.7 GHz computer with 8 GB of RAM was used for all of the experiments. Compared with the similar pose estimation study using SVM in the literature [124] whose registration time is 375.3 seconds and training time is 473.2 seconds, our method has an obvious advantage in the pose estimation speed. The training

of SVM models with our method requires a longer time [124], but it is an off-line process thus does not impact the performance of the on-line pose estimation. The training speed can be improved by parallelizing the training process in the future.

5.3.6 Discussions

We developed a pose estimation method as an initialization for 2D-3D registration. More precisely, an ASM based segmentation method was applied to extract the 2D contour from x-ray images. Normalized elliptical Fourier descriptors were used to represent the shape of the 2D contour. Then a hybrid classifier integrating KNN and SVM estimated the pose of the 3D model. Our experimental results demonstrated the reliability of the proposed initialization method for 2D-3D registration.

Automatic initialization is necessary for 2D-3D registration because different initial guesses do not generate a unique solution for the knee position using the automatic 2D-3D registration method [98]. The proposed method generates an accurate automatic estimation of the initial pose, leading to improved accuracy in the following 2D-3D registration. Moreover, partial occlusion occurs in fluoroscopic images due to the overlapping between femur and tibia. The proposed method is robust to the occlusion by segmenting the fluoroscopic images based on the data from the statistical atlas. Our method was applied to 2D-3D registration of femur, tibia, femur implant, and tibia implant in this paper, but it can be easily extended for various medical objects and imaging modalities for both monoscopic and stereoscopic registration. Reliable 2D-3D registration opens up many exciting possibilities such as pre-operative planning, intra-operative navigation, and diagnosing purposes by comparing different sources of data such as CT, MRI and X-ray fluoroscopy

images where the visualization of the anatomy is fundamental.

The limitation to the initialization method is that its resolution is limited by the increments of the training data (1 degree in each direction). Therefore this method is only suitable for initialization purpose. However, it can search in a large range of possible solutions, leading to the capture range as large as the domain of the training data. Moreover, the proposed initialization method does not need any information from 3D model during testing. Another limitation is related to mirror and circular symmetry, which leads to similar 2D projection images at symmetric views. Currently the proposed method doesn't address such cases, though this can be solved by including contextual information about the motion of the adjacent elements of the joint to distinguish between symmetric poses. The reason for the choice of angle range in this paper is to cover the domain of the clinical problem being addressed in the current application. It is not method dependent and can be easily expanded to larger ranges for different applications. The training time is also an issue, which can be improved by parallelizing the training process.

CHAPTER VI

3D REONSTRUCTION FROM MONOPLANE FLUOROSCOPY

Knee kinematics is of interest to a wide range of medical activities such as knee motion characterization, knee arthroplasty design, and knee disease diagnose. In clinical practice, it is desired to keep the number of X-ray fluoroscopic images to a minimum due to cost, acquisition and computation times and due to less constraint on patient's motion. Therefore, a 3D reconstruction method is proposed for knee kinematics analysis from monoplane fluoroscopy. The 3D model is generated based on a nonlinear statistical shape model. To increase robustness, a hybrid energy function is employed to combine the information from feature and intensity to measure the similarity between 3D model and x-ray image. The hybrid energy function requires neither time consuming DRR generation nor accurate segmentation which is error prone. Section 6.1 is the introduction, Section 6.2 describes the detailed approach, and Section 6.3 shows the experiments and results.

6.1 Introduction

Kinematics analysis has a wide application such as to analyze anatomical and functional characteristics of joint [4, 102], to diagnose diseases [125], and to characterize knee motion [126].

Monoplane X-ray fluoroscopy is a useful tool for analyzing joint kinematics in vivo [75, 77] because it allows sufficiently unconstrained motion of the patients, such as deep knee bending process. Bi-planar fluoroscopy using two orthogonal units, though may lead to more accurate results, would unacceptably constrain the motion of the patient. For this

purpose, monoplane fluoroscopy is used in this research to allow patient to move freely without impaired activities.

The key step in knee kinematics tracking is 2D-3D registration, which determines the shape and pose parameters of the 3D model in the camera coordinate system. Early researchers track the implants on fluoroscopic sequences through library based registration [75, 106]. Such tracking methods utilize the shape of the implants as known from the manufacturer. Moreover, by comparing the test case with a template library, such methods an exhaustive computation time and their accuracy is determined by the resolution of the template library. Other researchers developed rigid registration method to apply 2D-3D registration to natural bones [102, 127-129]. The 3D model is generated by segmenting 3D magnetic resonance (MR) [129] or computed tomography (CT) images [102, 127, 128]. Though sufficient accuracy can be achieved in these methods, bone segmentation can be labor intensive. Moreover, prior CT scan is required in these methods which exposes the subjects to additional radiation.

Reconstruction of bone shape from x-ray images is beneficial because it does not require prior CT or MR scan thus lower the radiation dose and acquisition costs. Statistical Shape Model (SSM) was first introduced by Cootes in 1995 as shape prior to restrict the reconstruction to plausible shapes [3]. The linear SSM is learned by a Principal Component Analysis (PCA) of the training shapes, which implies the assumption that the shapes are Gaussian distributed. While this model has been successfully applied to the 3D reconstruction of various structures in medical imaging, there are cases when the set of training shapes exhibits highly non-linear shape deformations, such as large shape variations between bones.

KPCA has been the focus of research in the pattern recognition community [130, 131]. The basic idea behind these methods is to map the data from an input space to a high dimensional feature space via some nonlinear map, and then apply a linear PCA in the feature space. After linear analysis in the feature, a reverse mapping back from feature space to input space is also non-trivial, because the exact pre-image typically does not exist and one can only settle for an approximation solution [130]. For certain invertible kernels, this nonlinear problem can be solved by a fixed-point iteration [130]. However, this method is depend on the initial starting point and is highly sensitive to local minima. To solve this problem, Kwok [131] developed a closed-form solution that directly finds the pre-image based on distance constraints in the feature space. It is non-iterative, involves only linear algebra and does not suffer from local minima. To the best of our knowledge, these approaches have never been applied to the kinematics tracking of 3D models in medical imaging field.

According to the type of similarity measure, 3D reconstruction methods can be categorized into intensity based and feature based registration. Intensity based registration compared the 3D model projection directly with the intensity information from 2D x-ray images. The registration is estimated by iteratively minimizing the similarity measure between a simulated x-ray projection images (digitally reconstructed radiograph (DRR)) from current transformation of the 3D images and the x-ray image [53, 54]. However, intensity-based registration may be sensitive to intensity changes in the fluoroscopic images by the overlapping of the other leg or surgical tool. Moreover, the generation of DRR is time consuming. Feature-based registration requires extraction of features in the x-ray image [6, 7, 21, 52]. Features can be points such as landmarks and contours extracted

from 2D images. Feature-based method is fast because it compared only part of the image and no time consuming DRR is required. However, the accuracy of this method depends on the accuracy of both segmentation of X-ray images and the determination of point correspondence, which may introduce errors to the registration process. We propose a hybrid method that integrate feature and intensity based method so that it is less time consuming than the intensity based method and less prone to segmentation error than feature based method. Unlike feature based method, we do not require an accurate segmentation of the knee bone silhouette in the image. Instead, we use a direct image-to-image similarity measure, which has been used in our previous work on 2D-3D registration of knee implant [5]. This study is different from [5] in that we employ prior knowledge to remove the outliers in the edge image to avoid impacts from noises and neighboring bones. For the intense information, we match the x-ray image directly with projected 3D model that is generated by 3D rendering. Unlike intensity based method, we do not use the time consuming DRR as simulated x-ray projection image. Instead, 3D model is rendered on the 2D plane by a software developed by the authors using a 3D graphic library.

6.2 Approach

The goal of this research is to reconstruct 3D model of knee bones from a sequence of single plane fluoroscopic X-ray images. The 3D model is built from nonlinear statistical shape model described in Section 6.2.1. Section 6.2.2 shows an initialization method that find the closest shape to start the optimization process. Section 6.2.3 is the optimization framework, and Section 6.2.4 describes the design of a novel energy function.

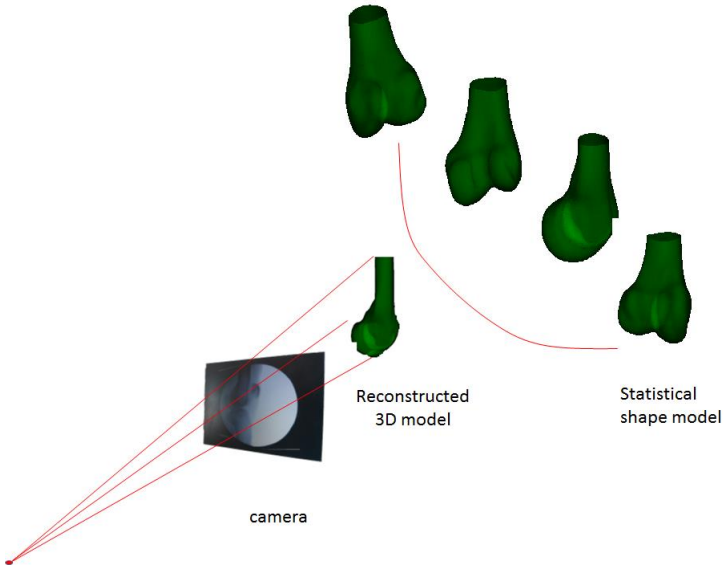


Figure 33. 3D reconstruction through nonlinear statistical shape model

6.2.1 Statistical Shape Model

A). SSM Training by Kernel Principal Component Analysis

SSM represents the global shape constraints with the dominant shape variation in the training set. The training data, $X \in \Re^3$, are 3D surface mesh models defined as $X = \left\{ X^\alpha \mid \alpha \in \{1,...,N\} \right\}$, and each training shape $X^\alpha \subset X$ has a set of vertices, where $X^\alpha = \left\{ x_n^\alpha \mid n \in \{1,...,M\} \right\}$, and where $x_n^\alpha = (x, y, z)$ denotes the coordinates of the nth vertex X^α. Corresponding points between each training shape are found before building the shape model with a statistical atlas [90].

The input data is mapped onto high dimensional feature space H via the nonlinear map $\Phi : \Re^3 \to H$. This map does not need to be explicitly known. Alternatively, one can introduce the kernel matrix, which is defined to be a function $K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$

such that for all data points x_i , $K = \begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \cdots & k(x_N, x_N) \end{pmatrix}$

is symmetric and positive semi-defined.

We solve the eigendecomposition problem

$$CV = \lambda V \tag{6-2}$$

where $C = \frac{1}{N}\sum_{i=1}^{N} \Phi(x_i)\Phi(x_i)^T$. Since the eigenvectors V lies in the span of $\Phi(x_i), \cdots, \Phi(x_N)$, equation (6-3) is equivalent to

$$\Phi(x_i) \cdot \bar{C}V = \lambda(\Phi(x_i) \cdot V) \text{ for all } i = 1,....,N \tag{6-3}$$

Since $K(x_i, x_j) = \left\langle \Phi(x_i), \Phi(x_j) \right\rangle$, and since K is symmetric, it has a set of eigenvectors which spans the whole space, thus this is equivalent to solve the dual eigenvalue problem

$$K\alpha = N\lambda\alpha \tag{6-4}$$

where λ is eigenvalue for both of the problems. Then standard principal component analysis is conducted.

The eigenvectors $\alpha^1, \cdots \alpha^N$ is normalized by requiring that the corresponding vectors in F be normalized, i.e. $\left\langle V^k, V^k \right\rangle = 1$ for all $k = 1, \cdots, N$. This is equivalent to

$$\begin{aligned} 1 &= \sum_{i,j=1}^{N} \alpha_i^k \alpha_j^k \Phi(X_i) \cdot \Phi(X_j) \\ &= \lambda_k(\alpha^k \cdot \alpha^k) \end{aligned} \tag{6-5}$$

Project the input data $X \in \Re^3$ on to kernel PCA space, spanned by the first l Eigenvectors of C is then given by

$$P^l \Phi(X) = \left\langle V^T, \Phi(X) \right\rangle = \sum_{i=1}^{l} \alpha_i \Phi(X_i) \cdot \Phi(X) = \sum_{i=1}^{l} \alpha_i K(X_i, X) \tag{6-6}$$

The assumption we made about $\Phi(x_i)$ is that $\Phi(x_i)$ is centered. We can relax this assumption by $\Phi(x_i) = \Phi(x_i) - \frac{1}{N}\sum_{i=1}^{N} \Phi(x_i)$, the corresponding kernel matrix is modified as follows

$$\bar{K}_{ij} = K - \frac{1}{N}\sum_j K_{ij} - \frac{1}{N}\sum_i K_{ij} + \frac{1}{N^2}\sum_i \sum_j K_{ij} \tag{6-7}$$

The detailed procedure is described in Algorithm 3.

Algorithm 3 Training KPCA model

Input: training data x

Output: kernel matrix K, α, eigenvalue λ, eigenvector α

1: construct the kernel matrix K with Gaussian kernel

2: center K with $\bar{K}_{ij} = K - \frac{1}{N}\sum_j K_{ij} - \frac{1}{N}\sum_i K_{ij} + \frac{1}{N^2}\sum_i \sum_j K_{ij}$

3: perform SVD on K as $K\alpha = N\lambda\alpha$

4: normalize the eigenvectors $\alpha^1, \cdots \alpha^N$ by $\frac{1}{\sqrt{\lambda_k}}$

B). Pre-image Approximation

Although the Φ map is not necessarily known, we would like to reconstruct the pre-image $\hat{X}$ of the corresponding test point $X \in \Re^3$ based on the distance constrain in the

input space. This is achieved by the relationship between input-space distance and feature-space distances [131].

For any two points x_i and x_j in the input space, we can obtain their Euclidean distance $d(x_i, x_j)$. Accordingly, we can also obtain their feature-space distance $\tilde{d}\left(\Phi(x_i), \Phi(x_j)\right) = 1 - \alpha K \alpha - 2K\alpha$ by the mapped images [132]. The relationship between $d(x_i, x_j)$ and $d\left(\Phi(x_i), \Phi(x_j)\right)$ can be derived as follows,

$$\tilde{d}^2\left(\Phi(x_i), \Phi(x_j)\right) = \left\|\Phi(x_i)\right\|^2 + \left\|\Phi(x_j)\right\|^2 - 2\Phi(x_i) \cdot \Phi(x_j) \tag{6-8}$$

For many commonly used kernels, there is a simple relationship between $d(x_i, x_j)$ and $d\left(\Phi(x_i), \Phi(x_j)\right)$. For Gaussian kernel of the form $k(\Phi_i, \Phi_j) = e^{-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}}$, which is invertible. The relationship can be described by

$$d_{ij}^2 = -2\sigma^2 \log\left(\frac{1}{2}\left(K_{ii} + K_{jj} - \tilde{d}_{ij}^2\right)\right) \tag{6-9}$$

According to the exponential impact of the neighboring points distance to the estimation of the current point according to the iterative scheme $\hat{x}_{t+1} = \dfrac{\sum_{i=1}^{N} \tilde{\gamma}_i \exp(-\left\|\hat{x}_t - x_i\right\|^2 / 2\sigma^2) x_i}{\sum_{i=1}^{N} \tilde{\gamma}_i \exp(-\left\|\hat{x}_t - x_i\right\|^2 / 2\sigma^2)}$ in [115], the contributions of x_i's drop rapidly with increasing distance from the pre-image, only *N* neighbors $\{x_1, \cdots x_N\} \in \Re^M$ are obtained. Then, a $M \times N$ matrix is constructed, $X = [x_1, \cdots x_N]$. Assuming that the training patterns span an M-dimensional space, we can obtain the singular value decomposition (SVD) of the centered $M \times N$ matrix XH as

$$XH = U\Lambda V' = UZ \tag{6-10}$$

where Z is a $M \times N$ matrix with columns z_i being the projection of x_i onto the e_j's. The squared distance of x_i to the origin, is equal to $\|z_i\|^2$ such that $d_0^2 = \left[\|z_1\|^2, \cdots, \|z_n\|^2\right]'$.

According to the assumption that the pre-image $\hat{x}$ is in the span of the n neighbors, its location will be obtained by requiring $d^2(\hat{x}, x_i)$ to be as close to those values in (6-9) as possible. Hence least square solution of $\hat{z}$ is yielded as follows

$$\hat{z} = -\frac{1}{2}\Lambda^{-1}V'\left(d^2 - d_0^2\right) \tag{6-11}$$

This $\hat{z}$ is expressed in terms of the coordinate system defined by the e_j. Transforming back to the original coordinate system in the input space, we thus have

$$\hat{x} = U\hat{z} + \bar{x} \tag{6-12}$$

The detailed procedure is described in Algorithm 4.

Algorithm 4 Pre-image approximation

Input: shape parameters θ, kernel matrix K, α, training data X

Output: reconstructed shape $\hat{x}$

1: calculate feature space distance by $\tilde{d}_{ij}^2 = 1 - \alpha K\alpha - 2K\alpha$

2: calculate input space distance by $d_{ij}^2 = -2\sigma^2 \log\left(\frac{1}{2}\left(K_{ii} + K_{jj} - \tilde{d}_{ij}^2\right)\right)$

3: select k nearest neighbor of the pre-image $\hat{x}$ by sorting $\tilde{d}_{ij}^2$

4: SVD on the k nearest neighbors by $XH = UZ$

5: estimate pre-image $\hat{z}$ by least square estimation as $\hat{z} = -\frac{1}{2}\Lambda^{-1}V'\left(d^2 - d_0^2\right)$

6: project back to original coordinate system in the input space by $\hat{x} = U\hat{z} + \bar{x}$

6.2.1 Find the Closest Shape

The two-step optimization framework discussed in Section 6.2.1 starts with the closest shape of the 3D model. It is determined by first encoding the training 3D shapes into shape parameters using KPCA. Then *N* cluster centers of the training shape parameters are found using K-mean clustering. Next, the 3D model reconstructed from the *N* clustering centers are projected onto 2D plane as binary images with known poses. The shape in the binary images are encoded with Fourier Descriptors. The closest shape to the object of interest in the testing X-ray image is determined by comparing with the training data using nearest neighbor. The flow chart is shown in Fig. 34.

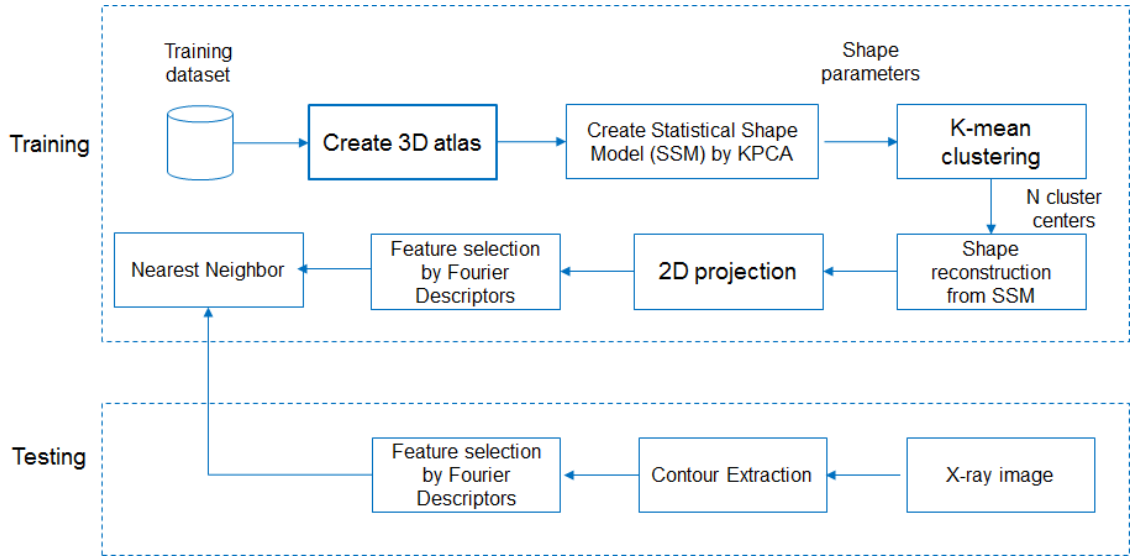


Figure 34. Flow chart of finding closest shape

6.2.2 Optimization

Let $S \in \Re^3$ be the smooth surface of the object of interest, and denote $X = [x, y, z]$, $X \in S$ to be the spatial coordinate that are measured with respect to

referential frame of the imaging camera. Let $X_0 \in \Re^3$ and $S_0 \in \Re^3$ be the coordinates and surface in the 3D world, respectively. Moreover, we assume a camera realization $\pi : \Re^3 \to \Omega$ such that $X' = [x/z, y/z]$ and $\Omega \in \Re^2$ denotes the domain of the image *I*. Similarly, we can form the edge image of the 3D model as $C = \partial I$ where $C \in \Re^2$.

We can locate the S in the camera reference frame via the transformation *T* such that $S = T(S_0)$ and the corresponding pointwise expression

$$X = T(X_0) = RX_0 + t \tag{6-1}$$

where R is a rotation matrix $R = R_\gamma * R_\beta * R_\alpha$, $R_\alpha = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}$,

$R_\beta = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}$, $R_\gamma = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and *t* is a translation vector, $t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$

. X_0 is reconstructed from SSM with the shape parameters $\theta_1 = \{b_1, \cdots, b_n\}$ via $X_0 = f(\theta)$ using kernel PCA, which will be discussed in Section 6.2.2.

Therefore, two sets of parameters are to be determined so as to align the 3D model to the 2D X-ray image. One is the shape parameter $\theta_1 = \{b_1, \cdots, b_n\}$.The other is the pose parameters, $\theta_2 = \{\alpha, \beta, \gamma, t_x, t_y, t_z\}$. The optimal parameters are determined by minimization an energy function with a global optimization algorithm, pattern search (PS). The energy function will be defined in Section 6.2.3. PS attempts to minimize a function by comparing its value in a finite set of trial points at each iteration. As a direct search

method, PS can be used on functions that are not continuous or differentiable. Convergence for stationary points can be proved from arbitrary starting points [133].

To cope with the fitting of multiple parameters, we divided the optimization into two stages. The first stage provides a robust initialization to the reconstruction. The closest shape is rigidly registered to the 2D x-ray image resulting in a close to optimal pose. The second stage comprises both tracking and shape reconstruction. In this stage, both pose and shape parameters are optimized, resulting in one shape per sequence and one pose per-frame. An illustration of the two-stage optimization process is shown in Fig. 35.

Minimization of the energy function in (6-13) is performed with a global optimization algorithm named pattern search (PS). PS attempts to minimize a function by comparing its value in a finite set of trial points. As a direct search method, PS can be used on functions that are not continuous or differentiable. Convergence for stationary points can be proved from arbitrary starting points [133].

6.2.4 Energy Function

The energy function E is defined to quantify how well the projection of the 3D model with the current pose and shape fits the corresponding object in the underlying fluoroscopic image. Two input data are involved in the energy function: one is a 2D X-ray image defined as $I_1 : \Omega_1 \in \Re^2$, the other is projected 3D surface mesh model defined as $I_2 = \pi(X)$, $I_2 : \Omega_2 \in \Re^2$. Similarly, the corresponding edge image can be defined by $C_1 = \partial I_1$ and $C_2 = \partial I_2$ where $C_1, C_2 \in \Re^2$. Accordingly, the energy function is calculated as the sum of a region score term E_r, an edge score term E_e, and a homogeneity term E_h as shown in Equ. (6-13)

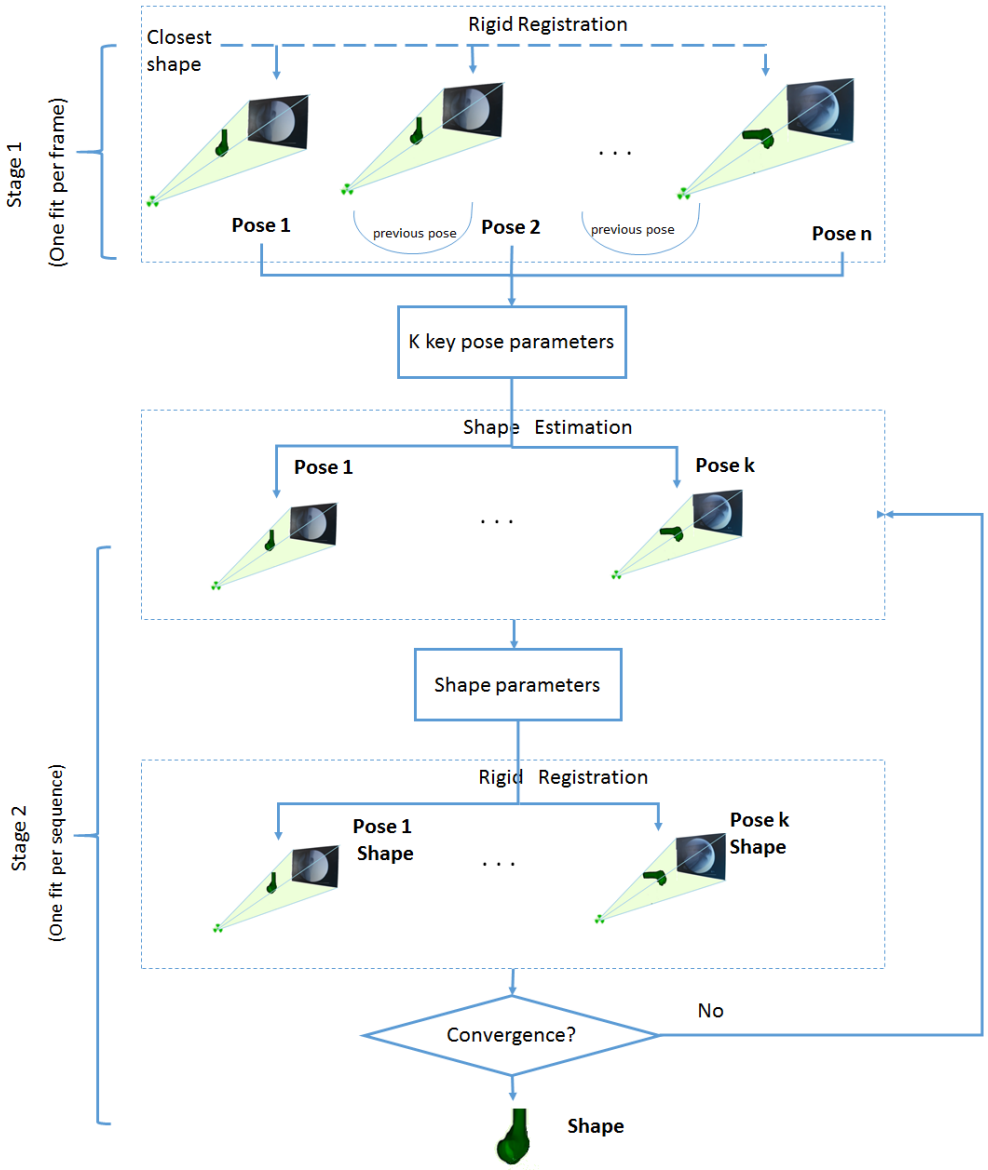


Figure 35. The two-stage sequence optimization. In the first stage the initial closest shape is rigidly aligned to the monoplane frames. k key poses are selected as input to the second stage optimization. The second stage alternates between shape and pose optimization until convergence occurs.

$$E = c_1 u(x) E_e + c_2 u(x) E_r + c_3 E_h \tag{6-13}$$

where c_i are weighting parameters that set the importance of each terms, $u(x)$ masks the images such that it results in 1 for pixels whose projection falls in the range of expanded initial 3D model projection and 0 otherwise.

E_e compares the edge feature in the two images. It is similar to a cross correlation between the two edge images[5] as follows

$$E_e = \frac{\sum_{i \in \Omega_1, j \in \Omega_2} C_1(i) C_2(j)}{\sum_{j \in \Omega_2} C_2(j)} \tag{6-14}$$

where C_1 is the edge image of 2D x-ray image, C_2 is the 3D model contour projected onto 2D plane. The edge score is a maximum when the projected 3D model contour coincides with the corresponding object edges in the 2D x-ray image. The generation of the edge image from 2D x-ray image is through a Prewitt filter plus masking the FOV area. An example of model and the corresponding edge image extracted from the fluoroscopic X-ray image is shown in Fig. 36.

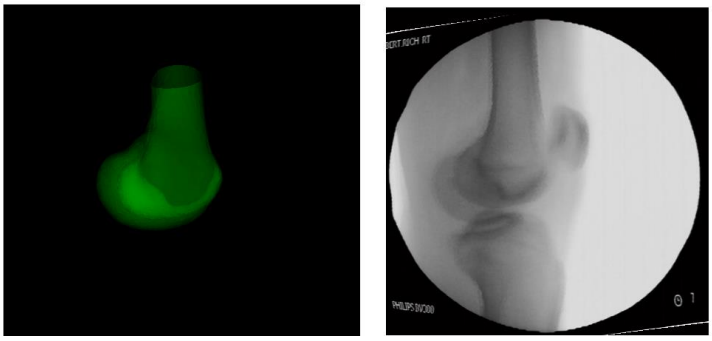
(a) (b)

Figure 36. Edge score, (a) is the model edge image, (b) is the edge image extracted from fluoroscopic X-ray image

E_r directly compares the intensity between I_1 and I_2. Local cross correlation (LCC) is used as similarity measure [134]. Our implementation is modified from [134], being able to handle local motion and intensity scaling. The detailed form of E_r is as follows

$$E_r\left(I_1, I_2\right) = \frac{1}{N}\sum_{j\in\Omega_1} \frac{\sum_{i\in n(j)} G_p(j)(I_1(i)-\bar{I}_1(j))^2 (I_2(i)-\bar{I}_2(j))^2}{\sum_{i\in n(j)} (I_1(i)-\bar{I}_1(j))^2 \sum_{i\in n(j)} (I_2(i)-\bar{I}_2(j))^2} \tag{6-15}$$

where I_1 is 2D x-ray image, I_2 is the 3D model projected onto 2D plane, $\bar{I}_1$ and $\bar{I}_2$ are the corresponding mean values, *i* represents a pixel in the neighborhood n(*j*) around pixel *j* in the 2D x-ray image. G_p is a Gaussian window function centered around each pixel in the 2D image. An example of model region image and fluoroscopic X-ray image is shown in Fig. 37.



(a) (b)

Figure 37. Region score. (a) is the model region image, (b) is the fluoro region image

The original cross correlation is frequently used for single-modality registration, where a linear relationship between the intensity of images can be assumed [135]. For multi-modality images, however, the linear relationship only holds in the small neighborhoods of the images [134]. Thus, a local cross correlation (LCC) is defined to

compute the similarity value within the neighborhood $n(j)$ of a number N of pixels in Ω_1 . Note from (2) that if one of the images has a constant intensity within $n(j)$ the contribution to LCC at j is zero. This implies that most contributions of LCC originate from edges in the images. The definition of LCC ensures $0 \le E_r\left(I_1, I_2\right) \le 1$. The better the structures in the neighborhood match, the larger the LCC value. Optimally, the well aligned images have values close to one.

Introducing a discrete, normalized Gaussian kernel, G_p , is based on two theoretical needs. The first one is the isotropy of the window, where only the distance from the point of interest influence the result. The second one is the effect of the graduate decrease of the influence of remote points. Remote points should have less influence than points located near the center of the window. Therefore, a discrete, normalized Gaussian kernel is chosen to meet the above requirements.

E_h is a homogeneity term defined as follows,

$$E_h = \sum\nolimits_{C_{in}} \frac{\left|I_1 - \bar{I}_{in}\right|}{n_{in}} + \sum\nolimits_{C_{out}} \frac{u(x)\left|I_1 - \bar{I}_{out}\right|}{n_{out}} \tag{6-16}$$

where the first term designates the variance of the set of intensity located on the internal of the femur 3D model projection; and the second term designates that located on the external of the femur 3D model projection. $\bar{I}_{in}$ is the mean value inside the 3D model n_{in} is the number of pixels inside the 3D model and $\bar{I}_{out}$ is the mean value outside the 3D model n_{out} is the number of pixels outside the 3D. $u(x)$ masks the images such that it results in 1 for pixels whose projection falls in the range of expanded initial 3D model projection and 0 otherwise. This energy term is minimal when the external contour of the

femur model projection delineates two homogeneous region separated by a model edge.

A plot of each term and the overall energy function is shown in Fig. 38. As we can see from Fig. 38(a) that there are a lot of local minimal for the edge score. This is due to the multiple match poses of the 3D model and the edge image. The curve for region score in Fig. 38(b) is smoother. Observe there is a saddle curve along the y translation axis. The peak of the saddle curve occurs where most overlay areas are dark in the model region image even with relative large overlapping area. The valley of the saddle curve occurs where most overlay areas are bright in the model region image even with relative small overlapping area. The homogeneity score in Fig. 38 (c) has a sharp global minimal which corresponds to the match between the 3D model projection and the X-ray image. Fig. 38(d) is the overall energy function. With the combination of three terms, the overall energy function is relatively smooth and has an obvious global minimal.

6.3 Experiments

Experiments were designed to evaluate from different aspects of the proposed method. The KPCA model behavior was evaluated. The reconstruction and tracking performance of the entire deep knee bending process was conducted.

6.3.1 Data

X-ray fluoroscopic images were acquired using a high-frequency video fluoroscopy unit. As the largest motion of the knee occurs in the flexion and antero-posterior translation, fluoroscopic imaging was conducted in the sagittal plane. The modern fluoroscopic systems allow video capture at a rate as high as 60 Hz. The fluoroscope is modeled as a perspective projection image formation model, which treats the fluoroscope sensor as

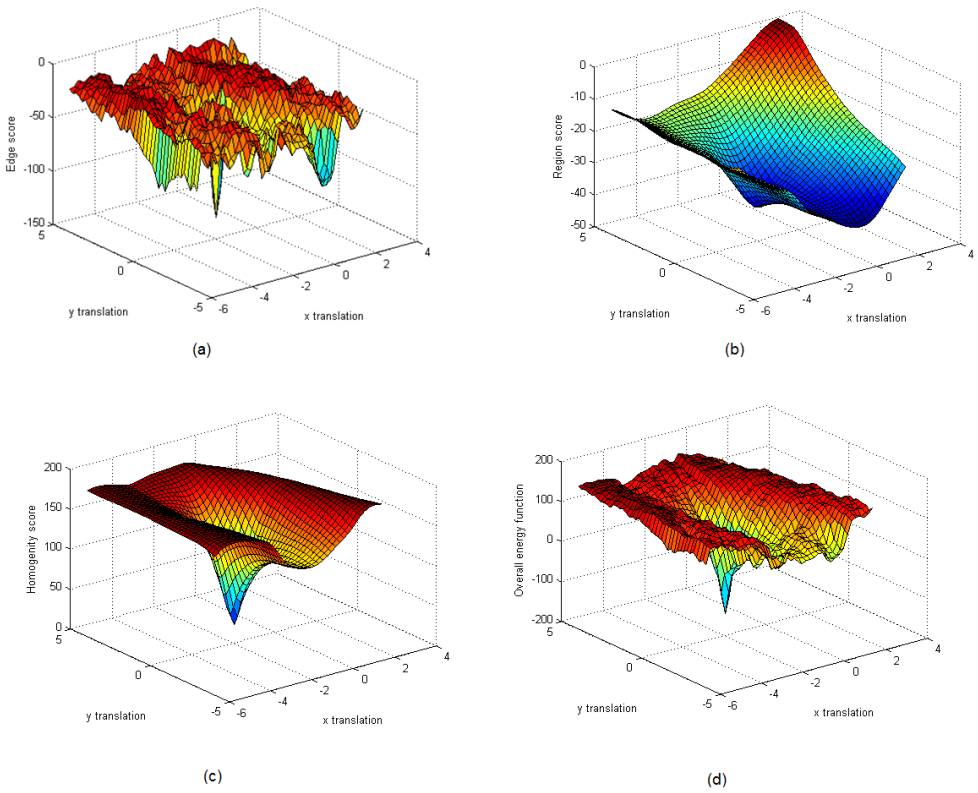


Figure 38. Energy function versus x and y axis translation. (a) edge score; (b) region score; (c) homogeneity score; (d) overall score. x axis is to the left, y axis is to the top, z axis is out of screen.

consisting of an X-ray point source and a planar phosphor screen upon which the image is formed. The x-ray images had a resolution of 640×480 pixels.

3D surface mesh model was obtained from a computer generated atlas of 258 three-dimensional surface mesh models. The atlas creation process involves adaptation of a template mesh to accurately match an input training model to be added to atlas. This adaptation will generate accurate dense points and surface correspondence [76].

6.3.2 Experiment Design

We performed two sets of experiments to evaluate different aspects of the proposed method. The first set of experiments evaluated the KPCA model behavior in terms of compactness, generalization, and specification. The results were compared with the performance of PCA to show the benefits brought by the nonlinear method. The training data are the 258 three-dimensional surface mesh model, and testing is conducted in a leave-one-out manner. RBF kernel is used as kernel, which has been a popular choice in the machine learning community and has proven to nicely extract nonlinear structures from data sets [136]. We determine the kernel parameter, σ, by trading off its compactness, generalization, and specification.

The second set of experiments shows the fit accuracy and tracking performance using the statistical shape model. The tracking was performed on five subjects of deep knee bending sequences. The starting poses were obtained by the automatic initialization method in [137]. The energy function has a factor λ, which determines the relative weight of each term. The value of λ is determined experimentally and λ_3 is negative because they are regulating terms to penalize the inhomogeneity and deviation from the mean shape.

6.3.3 Evaluation Scheme

A. Model evaluation

We compare KPCA with linear PCA to evaluate the statistical behavior of the nonlinear statistical model, in terms of compactness, generalization, and specificity.

The compactness of the model measures how few shape parameters are required for SSM to accurately describe the training data set. It is defined as the accumulative

variance $C(n_m) = \frac{\sum_{i=0}^{n_m-1} \lambda_i}{\sum_{i=0}^{N-1} \lambda_i}$, where $\lambda_i = \sigma_i^2$ is the i^{th} eigenvalue. $C(M)$ is measured as a function of the number of shape parameters M. Furthermore, the standard error of $C(M)$ is determined from the training size m: $\sigma_{C(M)} = \sum_{i=0}^{M-1} \sqrt{2/m}\lambda_i$.

Model generalization measures the ability of a SSM to describe instances outside of the training set. More specifically, the generalization of a SSM is measured using leave-one-out reconstruction. A SSM is built using all but one member of the training set and then fitted to the leaved-out example. The accuracy to which the model can reconstruct the leave-out example is measured by a function G(M) of the number of shape parameters M.

Model specificity measures the ability of a SSM to represent valid instances of the object class. It is defined as the average distance of uniformly distributed, randomly generated objects in the model shape space to their nearest member of the training set.

B. Evaluation of the 3D reconstruction method

To measure the accuracy of our approach, we used a distance error metric, namely root mean square (RMS) error. Let A and B be two surfaces to be compared, and supposing they are represented as point sets i.e, $A = \{a_0, a_1, \dots, a_{n-1}\}$ and $B = \{b_0, b_1, \dots, b_{n-1}\}$, we define RMS distance as follows,

$$RMS(A, B) = \sqrt{\frac{\sum_{i=0}^{n-1}(a_i - b_i)^2}{n}} \quad (6\text{-}18)$$

6.4 Results

We investigated the behavior of KPCA for five different values of the kernel parameter (σ=0.1, 1, 10, 25, 100) and compared them with the behavior of PCA. The shape

variance versus the number of modes used in statistical shape model for PCA and KPCA is shown is Fig. 39 to represent the compactness of the model. As we can see from the figure, with the increase of σ, the compactness of KPCA increases and with the decrease of σ, it is more linear. For σ=25, the compactness of KPCA exceeds PCA.

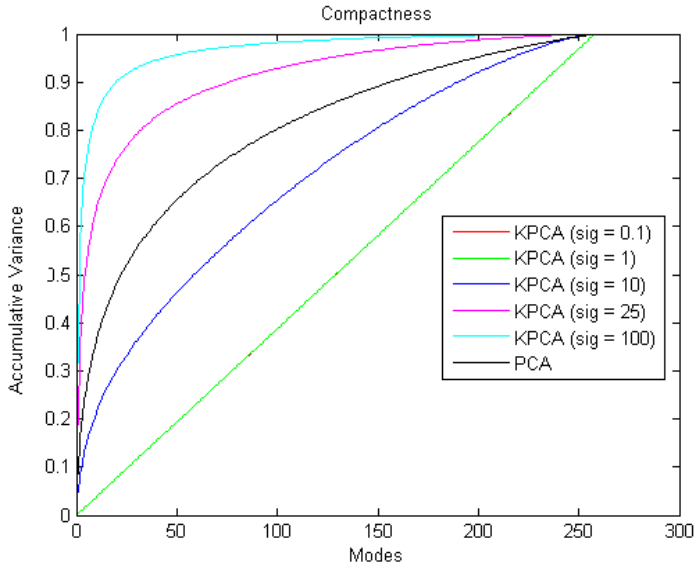


Figure 39. Compactness of the statistical models. Percentage of total shape variance versus the number of modes for statistical shape model using PCA and KPCA.

For generalization, we used the leave-one-out experiment to reconstruct the new shape with all but one sample. This is repeated for randomly selected six training samples. Finally, the average reconstruction error together with the standard deviation over the leave-one-out experiments was computed. Fig. 40 shows the reconstruction error for KPCA and PCA as a function of the number of variation modes. The reconstruction error of KPCA is a little higher than that of PCA. RMS error decreases with the increase value of σ; and the RMS error decreases with the increase number of modes involved.

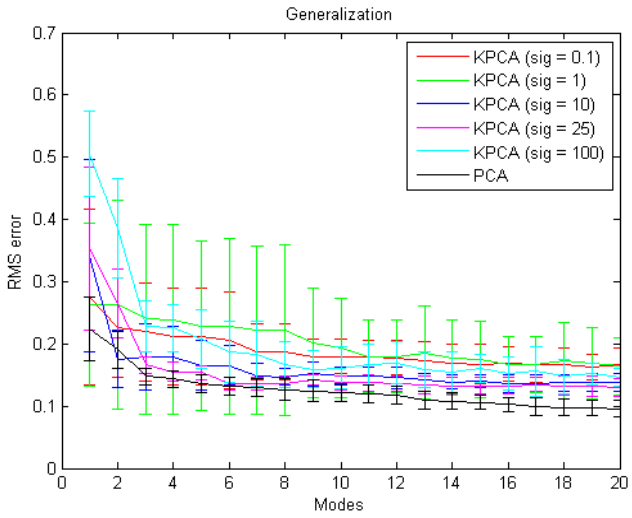


Figure 40. Generalization of the statistical models.

The specificity of the model is shown in Fig. 41, where (a) shows all the cases and (b) excludes the case of σ=100. As can be seen from the figure, the reconstruction error of KPCA is smaller than that of PCA for most values of σ . The specification of KPCA in the case of σ=1 and σ=10 outperforms PCA; whole σ=25 outperforms PCA in most of the variation modes.

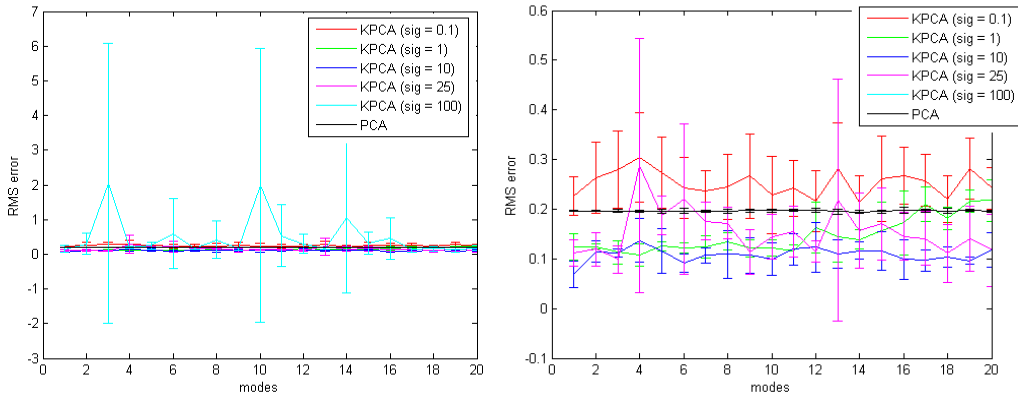


Figure 41. Specificity of the statistical models. (a) shows all the cases and (b) excludes the case of σ=100.

To demonstrate the ability of the proposed nonlinear model for 3D reconstruction and tracking from monoplane fluoroscopic image, video fitting was conducted for five subjects in a large sequence of closely spaced 2D x-ray fluoroscopic images (around 600 fluoroscopic x-ray images) during deep knee bending (DKB). The 3D model was fitted to the 2D fluoroscopic x-ray image frame by frame with the fitted pose of the previous frame as initialization of the next frame. Accuracy is calculated by comparing the projected 3D model contour on the 2D plane with the manually segmented 2D x-ray image. The results in shown in Table 6. Fig. 42 shows femur and tibia 3D model edge overlay on the x-ray images in key poses during deep knee bending process.

Table 6 Reconstruction accuracy

Frame #	Femur error (mm)	Tibia error (mm)
1	1.0732 ±0.5086	1.3320 ±0.2612
2	1.1815 ±0.3757	1.1252 ±0.4279
3	1.1475 ±0.5146	1.0844 ±0.3057
4	1.1792 ± 0.2727	1.1313 ±0.4206
5	1.2145 ±0.4295	1.2665 ±0.3720
6	1.2613 ±0.1466	1.0471 ±0.2810
7	0.9662 ± 0.1842	1.1114 ±0.2714
8	1.2161 ±0.2241	1.3938 ±0.4017

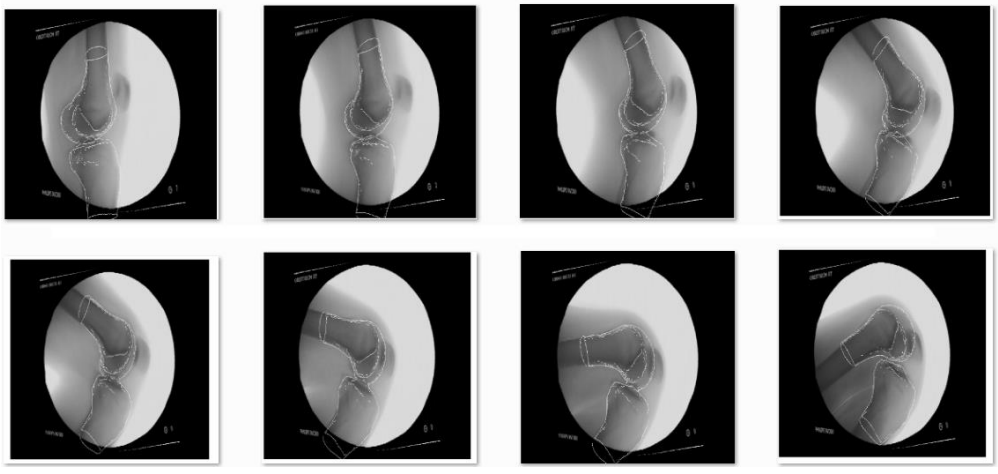


Figure 42. Femur and tibia 3D model edge overlay on the x-ray images in key poses during deep knee bending process

The RMS error of surface points femur reconstruction compared to the 3D model segmented from CT as gold standard for the five cases is 1.0713 ±0.0908 mm and that for tibia is 1.0453 ± 0.0694 mm. The distance map for femur and tibia is shown in Fig. 43 and Fig. 44, respectively. As shown in Fig. 43, large errors of femur 3D reconstruction occur in the areas which are not visible in 2D images, such as intercondyle fossa, medial condyle, and popliteal surface. The large error in medial condyle of femur is due to blur in the X-ray image. As shown in Fig. 44, the large error of tibia occurs in tibia tuberosity and edges on its two condyles, which is due to occlusion from the different part of the tibia.

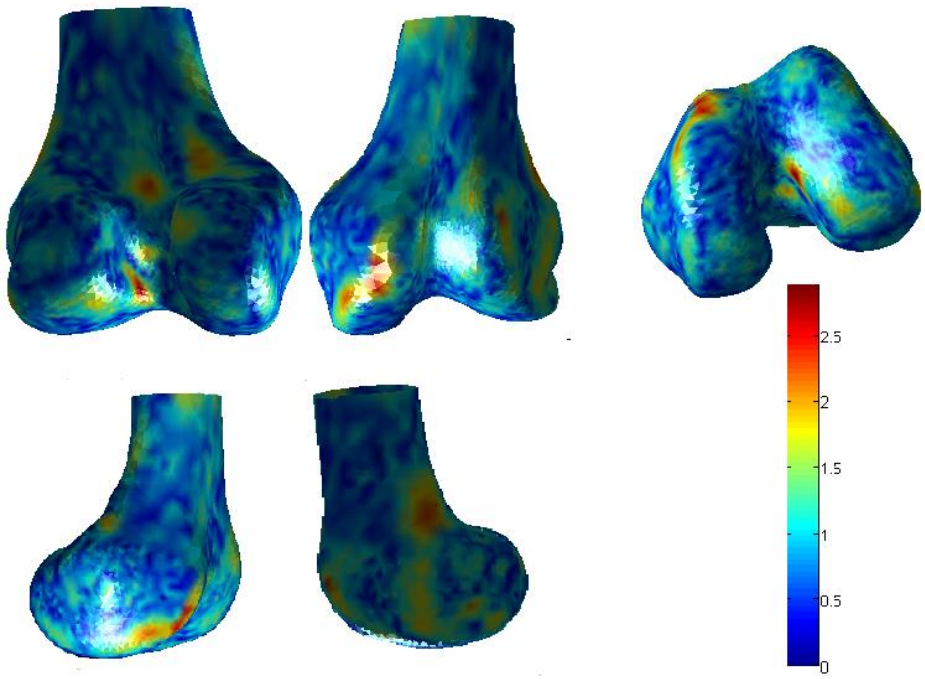


Figure 43. Femur surface distance map between reconstructed 3D model and the CT model

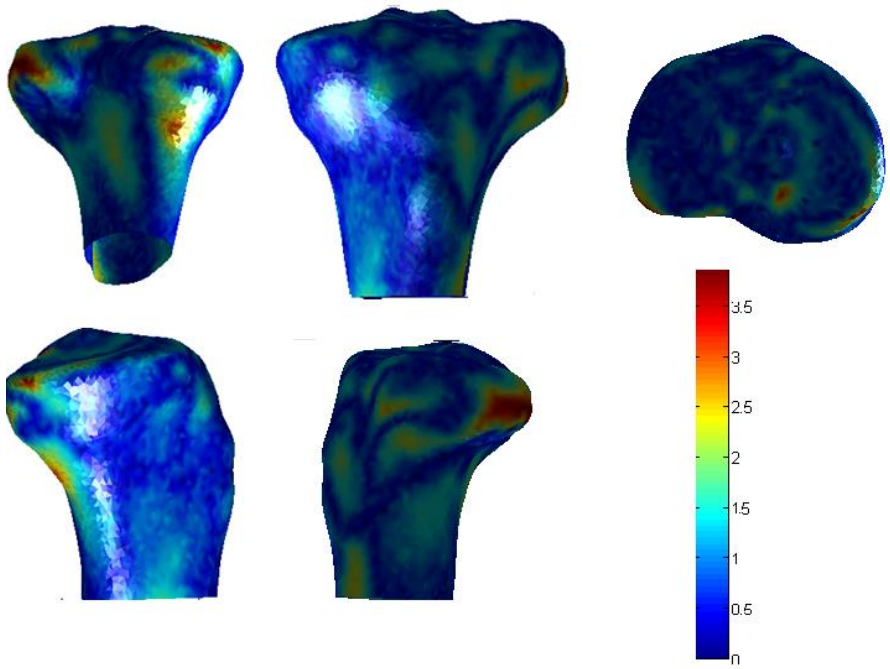


Figure 44. Tibia surface distance map between reconstructed 3D model and the CT model

CHAPTER VII

KNEE KINEMATICS ANALYSIS

Knee kinematics is of interest to a wide range of medical activities such as knee motion characterization, knee arthroplasty design, and knee disease diagnose. In clinical practice, it is desired to keep the number of X-ray fluoroscopic images to a minimum due to cost, acquisition and computation times and due to less constraint on patient's motion. Therefore, we proposed a 2D-3D registration method for knee kinematics analysis from monoplane fluoroscopy.

7.1 Introduction

Kinematic knee measurements have been extensively used to characterize knee motion [126], enable better total knee arthroplasty design [138], and provide useful information for diagnosing purpose [125].

X-ray fluoroscopy has been shown to be a useful tool for analyzing joint kinematics in vivo [75, 77]. X-ray video fluoroscopy is a sequence of x-ray images, where metallic implants or bones appear much darker than the soft tissues surrounding them, allowing for direct observation and analysis of their movement under dynamic and weight bearing conditions. Moreover, fluoroscopy is noninvasive and relatively low risk to the patient. A typical one minute protocol gives the patient a radiation exposure on the order of 0.6 to 1.8 "rad equivalent man" (rem) [4]. Since we wish to measure kinematics during activities such as deep knee bending, the patient's movement must be sufficiently unconstrained. For this purpose, we use monoplane fluoroscopy to allow patient to move freely without impaired activities. Bi-planar fluoroscopy using two orthogonal units, though may lead to

more accurate results, would unacceptably constrain the motion of the patient. For this purpose, we use monoplane fluoroscopy as our imaging method.

The key step in knee kinematics tracking is 2D-3D registration, which align the 3D model to the 2D x-ray image. There are three approaches for 2D-3D registration.

The first method is a library based registration, which matches the object to a template library that is precomputed before registration [75, 106]. The pose of the object was determined by finding the silhouette of the object with the closest match in the library. These methods are intuitive and simple to be implemented. However, they have the disadvantage that the object's contour must be accurately segmented from the image to achieve high prediction accuracy. Moreover, the accuracy of these methods depends on the template resolution and fine resolution requires large number of template which is neither space nor time efficient. Therefore, library-based registration usually serves as a coarse registration for initialization.

The second method is feature-based method, which requires features in the 2D x-ray images to be identified in the images prior to their registration. Features can be points such as landmarks and contours extracted from 2D images. The simplest feature-based 2D/3D registrations are point-to-point registration of corresponding anatomical landmarks in both 2D and 3D images [11, 31, 38-40]. The landmarks are usually located by an operator, thus it is heavily dependent on the skills of the operator. This method often serves as a rough initialization after which a finer registration method is applied. The problem of user defined landmarks is avoided by curve-to-curve registration. When a projection strategy is applied, segmented 3D images are projected onto 2D plane to match the contours extracted from 2D x-ray images by minimizing the sum of distance between

corresponding curves [6, 7, 32, 41-43]. By the back-projection strategy, virtual rays from 2D points on the intra-interventional image are formed and the distance between back-projected points in 3D space and the corresponding points on the 3D images are minimized [9, 44-48]. Zheng et al. proposed a hybrid approach that combines the projection and reconstruction strategies. First surface points on the 3D image are projected onto 2D image to find the 2D correspondence. These corresponded 2D points are then back-projected to 3D space to match the 3D surface points. However, the accuracy of feature based method depends on the accuracy of both segmentation of x-ray images and the determination of point correspondence, which may introduce errors to the registration process.

The third method is intensity-based registration, which relies solely on the intensity information from 2D and 3D images, respectively. This method matches the 2D and 3D images by minimizing the sum of intensity difference between various modalities. In contrast to feature-based method, intensity-based registration does not need segmentation; moreover the correspondence in intensity-based registration is more explicit thus making it a more accurate method than feature-based registration. The registration is estimated by iteratively minimizing the similarity measure between a simulated x-ray projection images (digitally reconstructed radiograph (DRR)) from current transformation of the 3D images and the x-ray image. The frequently used similarity measures include cross correlation[10], entropy[53], mutual information[54], gradient correlation[58], pattern intensity[18]. However, intensive-based method may be sensitive to intensity changes in the fluoroscopic images by the overlapping of the other leg or surgical tool. Moreover, the generation of DRR is time consuming.

We propose a hybrid method that combines the information from feature, intensity, as well as dynamic prior knowledge to form the energy function. For the feature information, we use an edge score that matches the edges extracted from x-ray images. Unlike feature based method, we do not require an accurate segmentation of the knee bone silhouette in the image. Instead, we use a direct image-to-image similarity measure, which has been used in our previous work on 2D-3D registration of knee implant [5]. This study is different from [5] in that we employ prior knowledge to remove the outliers in the edge image to avoid impacts from noises and neighboring bones. For the intense information, we match the x-ray image with projected 3D model that is generated by rendering. Unlike intensity based method, we do not use the time consuming DRR as simulated x-ray projection image. Instead, 3D model is rendered on the 2D plane by a software developed by the authors [5] using a 3D graphic library. The similarity measure for intensity information is local cross correlation (LCC), which is to reduce the difference resulted from multi-modality of the images. Gaussian window is used to gradually decrease the influence of remote points from the center of the window. For the multi-object fitting process, information from neighboring bones in the same frame can also be utilized to improve the fitting process by adding a large penalty to collision and mis-alignment between neighboring bones.

7.2 Approach

The goal of the proposed method is to track the 3D model of knee bones from the monoplane fluoroscopic sequence. The 3D model is the surface mesh model segmented from CT image. The detailed generation process is discussed in Section 7.3.1. The outline

the general framework is shown in Figure 45. The pose of the 3D model for all the sequences are determined by optimization an energy function. The energy function consists of an edge score term, an intensity score term, a dynamic prior term, and multi-object registration term. The edge and intensity term show how well the projections of the 3D model matches the fluoroscopic image with respect to edges and intensity, respectively. The dynamic prior term and multi-object registration term serves as regulation terms that penalize large pose difference between neighboring poses and misalignment or collision between neighboring bones in the same frame. The details of each component of the energy function is discussed in Section 7.2.1. The optimization strategy for the entire sequence is discussed in Section 7.2.3.

7.2.1 Energy Function

There are two input data involved in the registration process: one is a 2D X-ray image defined as $I_1 : \Omega_1 \subset R^2$ and the other is the 3D surface mesh model extracted from CT image defined as $I_2 : \Omega_2 \subset R^3$. Let T be a finite dimensional group with parameters $\theta_1, \theta_2, \cdots, \theta_n$ which transforms 3D image from it coordinate system to 2D image coordinate

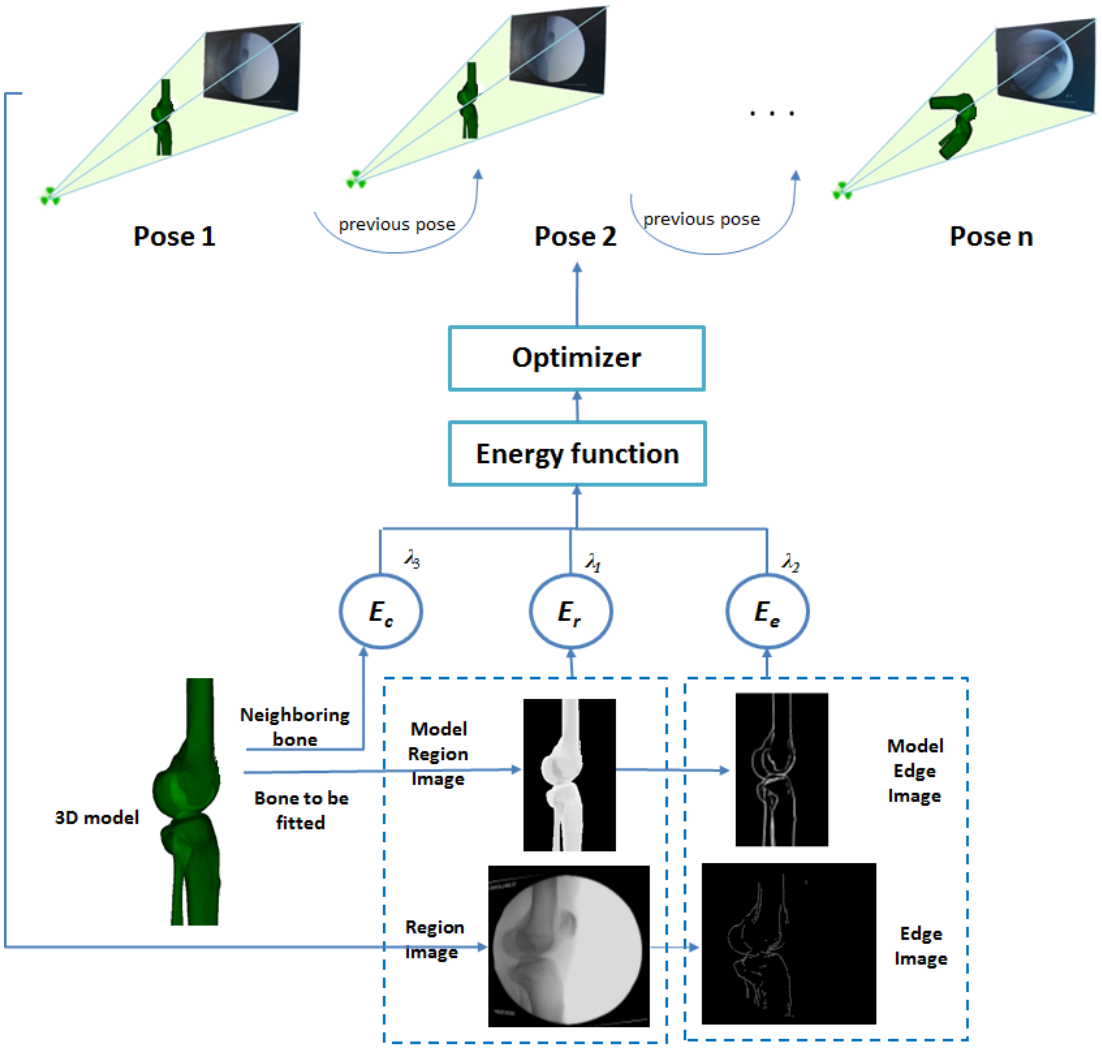


Figure 45. Method overview: The energy function is a linear combination of region score term E_r, an edge score term E_e, and a collision detection term E_c. The region score is defined by a normalized local cross correlation of model region image and the x-ray fluoroscopic image. The edge score is defined as a cross correlation of model edge image and edge image extracted from the original x-ray fluoroscopic image. The dynamic prior term uses information from the pose predicted in the previous frame as a regulation term. The multi-object registration term uses information from neighboring bones to avoid collision between bones and misalignment. The linear combination of the above four terms forms the energy function, which is then minimized to get the final estimation of pose of the 3D model.

system. We will denote by $x_1 \in \Omega_1$ and $x_2 \in \Omega_2$ the points in 2D and 3D images respectively. Our goal is to find a transform T under which $x_1 = T(x_2)$.

The proposed energy function E quantifies how well the projection of the 3D model with the current pose fits the corresponding object in the underlying fluoroscopic image, and includes two penalty terms. One is to avoid collision between neighboring bones in the same frame and the other is to keep the pose close to the one in the previous frame. Accordingly, it is calculated as the sum of a region score term E_r, an edge score term E_e, and a collision detection term E_c as shown below

$$E = -\lambda_1 E_r - \lambda_2 u(x) E_e + \lambda_3 E_c \tag{7-1}$$

where λ are weighting parameters; $u(x)$ masks the edges such that it results in 0 for edges whose projection falls in the range of neighboring bones or out of FOV in the x-ray fluoroscopy and 1 otherwise. E_r and E_e are same as the definition in (6-14) and (6-15). So, we will skip these two terms and discuss E_c only.

Due to the presence of neighboring bones and high noise or low resolution in the x-ray fluoroscopy, misregistration frequently occurs that is manifested as an overlapping of the neighboring bones. One way to improve the registration result is to account for spatial relations between the involved objects so that a simultaneous registration of all the involved bones are applied where the objects are interacting on each other through the registration. For this purpose, we added a collision detection term E_c for misalignment between bones in the same frame as defined below

$$E_c(\Theta_1, \Theta_2) = H_1(\Theta_1, \Theta_2) C_1 \tag{7-2}$$

where Θ_1 and Θ_2 are the pose parameters of the two involving bones; C_1 is weighting constants; H_1 is the Heaviside function , $H_1(\Theta_1,\Theta_2)=\begin{cases}1, & if \text{ collision occurred}\\ 0, & \text{else}\end{cases}$. It is used to penalize the overlapping between two surface mesh. If no collision is present, $H_1(\Theta_1,\Theta_2)$ will be equal to zero. For higher values of C_1, collision is more punished. It is time consuming to use standard collision detection library. An effective implementation of the collision detection is achieved by using the part of the meshes so that only adjacent surface patches of neighboring meshes is involved by the prior knowledge that collision between femur and tibia occurs only in femur condyle and tibia plateau and collision between femur and patella occurs in the femur trochlear groove. Therefore, search is done only in the surface meshes that potentially shows an intersection with a surface of a neighboring mesh so that computational complexity is significantly reduced. Note that E_c does not generally prohibit misalignment but makes points that lie inside other meshes or large distance between two bones in the out of screen direction less attractive.

7.2.2 Edge Image

It is crucial to generate the edge image that is free of noise, neighboring bones, and soft tissues. The image after Sobel filter may potentially contain irrelevant edges, most of which can be removed by prior knowledge, including edge length, and dynamic prior. The edge length feature includes the prior knowledge that noise edges tend to be short. All pixels of the edges length smaller than the threshold can be regarded as noise and removed from edge image. The dynamics prior uses the pose from previous frame due to the fact that high frequency acquisition of the fluoroscopic sequence makes neighboring poses

close to each other. Therefore, expanded previous neighboring bones are used as masks to remove the influence from neighboring bones.

An illustration of the edge image generation for femur is given in Fig. 46.

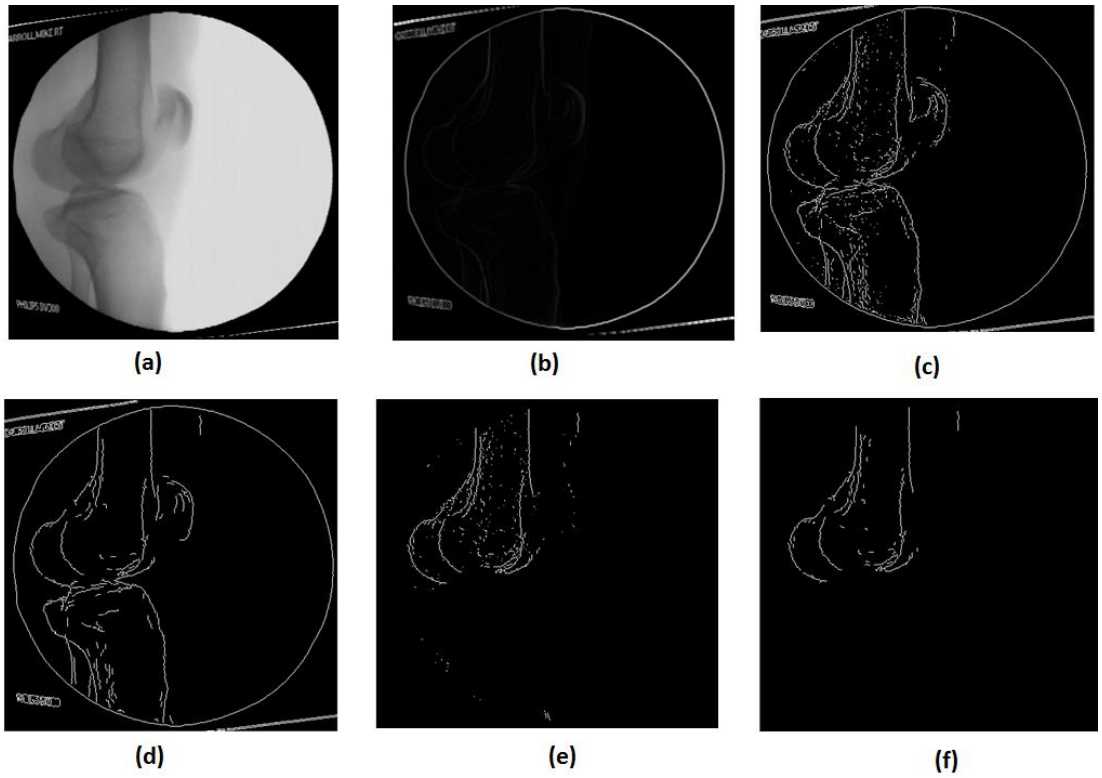


Figure 46. Edge image generation. (a) Fluoroscopic image; (b) gradient image; (c) edge image after thresholding; (d) edge image after removing small edges; (e) edge image after applying dynamic prior; (f) final edge image.

7.2.3 Sequence Optimization

Notice that the energy function (7-1) depends on the transformation matrix T, which is represented by a rotation matrix R and a translation vector t.

$$T(x_2) = Rx_2 + t \tag{7-3}$$

where $t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$, $R = {}_{\gamma} * R \quad {}_{\beta} *$ $R_\alpha = \begin{bmatrix} 1 & 0 & \\ 0 & \alpha c & -\mathrm{o}\ \alpha \\ 0 & \alpha s & \mathrm{i}\ \alpha \end{bmatrix}$;

$R_\beta = \begin{bmatrix} \mathrm{c} & \beta \mathrm{o} & \mathrm{s} & \beta \\ 0 & & 1 & \\ -\mathrm{s} & \beta \mathrm{i} & \mathrm{n} & \beta \end{bmatrix}$ $R_\gamma = \begin{bmatrix} \mathrm{c} & \gamma \mathrm{o} - & \mathrm{s}\gamma & \\ \mathrm{s} & \gamma \mathrm{i} & \mathrm{n}\gamma & \\ 0 & & 0 & \end{bmatrix}$. Thus the parameters to be optimized in the energy function (7-1) turns into $(t_x, t_y, t_z, \alpha, \beta, \gamma)$.

To cope with the fitting of multiple objects with continuous motions in a sequence, we divided the optimization into two stages.

We fit the multiple objects frame by frame with the previous frame's fitted pose as the initial pose of the next frame. For the first frame, the initial pose is set interactively by input the upper left and lower right point of the bounding box of each bone.

For each pose within one time frame, multiple objects are fitted in the sequence of femur, tibia, and patella. During fitting of each bone, 3D model of the rest bones are set as invisible resulting in only one bone shown in the model region and model edge image. However, the information of the other bones are used as dynamic prior for greater robustness fitting result. The details on dynamic prior can be found in Section 7.2.1 and Section 7.2.2.

Minimization of the energy function in (7-1) is performed with a global optimization algorithm named pattern search (PS). PS attempts to minimize a function by comparing, at each iteration, its value in a finite set of trial points. As a direct search method, PS can be used on functions that are not continuous or differentiable. Convergence for stationary points can be proved from arbitrary starting points [133]. Prior experiments

showed that the estimated parameters converge within thirty iterations. The energy in (7-1) is multiplied by -1 for minimization. Fig. 47 shows the energy function as a function of the number of iterations, for the same image fluoroscopy image. The upper dotted curve shows the instantaneous score at each iteration, as the optimization algorithm tests nearby model poses to see if the score will improve in various direction.

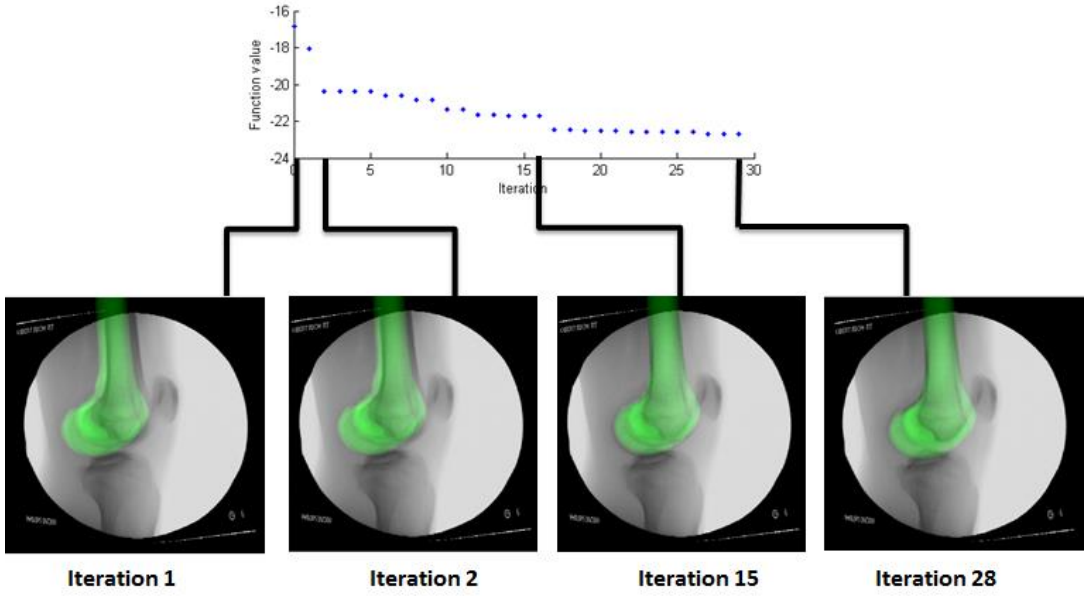


Figure 47. Energy function as a function of the number of iterations, on the same fluoroscopy image.

7.3 Experiment and Results

Experiments were designed to evaluate from different aspect of the proposed method. The tracking performance of the entire deep knee bending process was conducted. The fitting accuracy and dynamics was recorded for evaluation.

X-ray fluoroscopic images were acquired using a high-frequency video fluoroscopy unit. As the largest motion of the knee occurs in the flexion and antero-posterior translation,

fluoroscopic imaging was conducted in the sagittal plane. The modern fluoroscopic systems allow video capture at a rate as high as 60 Hz. The fluoroscope is modeled as a perspective projection image formation model, which treats the fluoroscope sensor as consisting of an X-ray point source and a planar phosphor screen upon which the image is formed. The x-ray images had a resolution of 640×480 pixels. The fluoroscopic system is shown in Fig. 48. 3D surface mesh model was obtained by the segmentation of Computed tomography (CT) images. Each bone was scanned with a $0.625 \times 0.625 \times 0.625$ mm^3 voxel setting. Sizing was consistent at 0.18%, compared with manual landmark measurements. Each bone was manually segmented, neglecting hyaline cartilage. Triangulated surface models were then generated using Amira software (Mercury Computer Systems, Chelmsford, Mass). The 3D models were then rendered on the 2D plane by a software developed by the authors [5] using a 3D graphic library (Open Inventor, TGS, San Diego, CA).

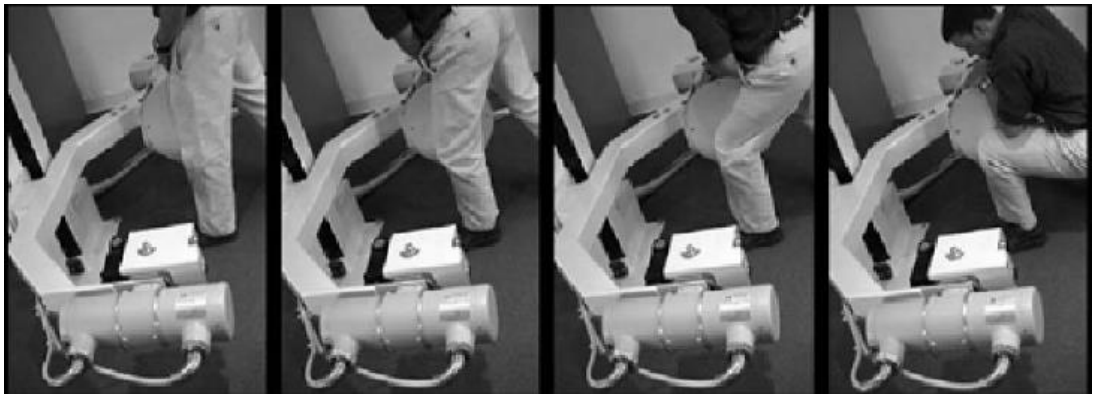


Figure 48. Monoplane fluoroscopic system for deep knee bending.

With monoplane fluoroscopy imaging, video fitting was conducted for five subjects in a large sequence of closely spaced 2D x-ray fluoroscopic images (around 600

fluoroscopic x-ray images) during deep knee bending (DKB). The 3D model was fitted to the 2D fluoroscopic x-ray image frame by frame with the fitted pose of the previous frame as initialization of the next frame.

To evaluate the performance of the proposed method, we used six frames from each sequence to represent the key pose in motion and compared the predicted poses in each frame with the ground truth pose. The ground truth pose is the manual fitting of the 3D model to the 2D images. We report in Table 8 the mean error with standard deviation of the motion differences between the ground truth and the predicted pose in terms of translation and rotation for femur and tibia. The model edge overlay on the X-ray images in the key poses during deep knee bending is shown in Fig. 49.

Table 7: Results of The Tracking With All Fits of The Five Test Sequences. Motion Errors Show The Mean Error with it Standard Deviation. The Unit for Translation Is in mm and for Rotation is in Degree

	Femur	Tibia
T_x (mm)	0.4822 ±0.9657	0.1056 ±1.3478
T_y (mm)	-0.3565 ±0.9365	-0.3504 ± 1.5945
T_z (mm)	-0.2035 ± 1.7815	0.6126 ±1.7868
R_x (°)	0.6554 ± 1.1047	-0.3310 ±1.1547
R_y (°)	0.7656 ±1.3567	-0.0898 ±1.0433
R_z (°)	0.5037 ± 0.8816	-0.1142 ±1.1305

Note from Table 7 that error of Z axis translation (around the medial-lateral axis) is significantly higher than that of other axis translations for all the three bones, because less information from the x-ray image is provided in that direction. The displacement in Z axis leads to scaling effect in the shape, which is much less sensitive to displacement in that direction. The most significant error for femur rotation is in Y axis (around the

proximal-distal axis). The bone surface have a relatively short distance to this axis, resulting in rotation around this axis causes only small changes in the contour shapes. Moreover, the shape variance of the condyles are relatively more than that of the shaft. A longer shaft therefore might down-weight the effect of rotation on the energy function. The major error occurs in the X axis (around the anterior-posterior axis) for tibia fitting because the rotation around this axis leads to least change in contour shape of tibia.

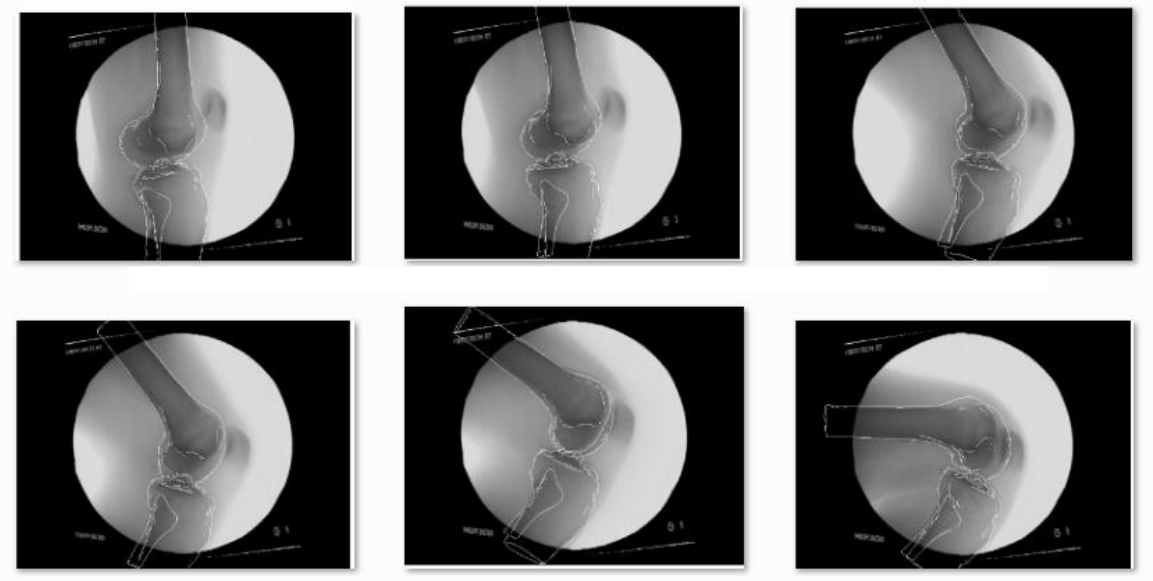


Figure 49. Model edge image overlay on X-ray images.

The tracking results shown in Table 7 is comparable to rigid registration results found in the literature, despite the lower quality fluoroscopic images and monoplane x-ray fluoroscopy used in this study. [6] reported an accuracy of -0.06 ± 0.63 mm in x-axis translation, 0.06 ± 0.58 mm in y axis translation, -0.06 ± 0.58 mm in z axis translation, 0.15 ± 0.86 degree in x axis rotation, 0.01 ± 0.69 degree in y axis rotation, -0.32 ± 1.23 degree in z axis rotation with biplane of femur. [19] reported a RMS error of 0.3 mm in x-axis

translation, 0.7 mm in y axis translation, 0.9 mm in z axis translation, 1.2 degree in x axis rotation, 0.6 degree in y axis rotation, 1.4 degree in z axis rotation for in-vitro dry femur registration with three fluoroscopic images. [48] reported an accuracy of -1.1578 ±1.4841 mm in x-axis translation, -0.8911 6±0.5634 mm in y axis translation, -0.2089 ± 0.9967 mm in z axis translation, -0.7189±0.5737 degree in x axis rotation, 0.1078±0.2704 degree in y axis rotation, -0.3733 ± 1.265 degree in z axis rotation for skull phantom data for biplane registration. Features are extracted manually in [48]. All the above methods use two or more x-ray images for registration. Also [48] and [19] use single time-point x-ray image for the registration. Care should be taken, though, when comparing the quantitative results directly, as different experimental conditions such as x-ray acquisition methods, test subjects, FOVs, etc are used in all studies.

Fig. 50 shows an example of a sample fitting result. The 3D model is on the left and the overlay of the edge of the 3D model on the x-ray image is on the right.

Figure 50. An example of a sample fitting result. (a) is 3D model overlay on the X-ray image; (b) is an overlay of projected 3D model edge image on the X-ray image.

CHAPTER VIII

CONCLUSIONS

In vivo kinematic analysis using x-ray video fluoroscopy is of value in a wide range of medical activities such as knee motion characterization, knee arthroplasty design, and knee disease diagnosis. X-ray video fluoroscopy captures a sequence of x-ray images of the patient in real time, allowing direct observation and analysis of the movement under dynamic and weight-bearing conditions. Moreover, fluoroscopy is noninvasive and relatively low risk to the patient. This study focuses on designing a practical, robust, and efficient system to reconstruct 3D model from monoplane fluoroscopic images with a nonlinear statistical shape model. Such reconstruction method is more time and cost efficient than the rigid 2D/3D registration methods, which require segmentation of CAD/CT/MRI for each patient to obtain the 3D model. In clinical practice, minimizing the number of X-ray fluoroscopic images is desirable due to lower cost, shorter acquisition and computation times, and less constraint on patient's motion. Reconstructing from monoplane fluoroscopic images allows the patient to move freely during kinematics analysis compared with previous kinematic works on 3D reconstruction which use two or more 2D images.

A robust segmentation method is necessary to extract features from 2D fluoroscopic x-ray images. In Chapter 4, a novel 2D segmentation method is presented based on active shape model. First the image is enhanced by a spectral clustering based method. Segmentation is then performed with ASM, leading to robust and accurate results.

Since the 2D-3D registration techniques are generally very sensitive to the initial alignment of the 3D model, an automatic initialization method is developed in Chapter 5 to estimate the 3D model's initial pose before 2D-3D registration. Contour of the knee bone or implant was first automatically extracted from 2D X-ray image. Shape descriptors were calculated by NEFD to represent the contour shape. The optimal pose was then determined by a hybrid classifier combining KNN and SVM. The proposed method can be easily utilized for 2D-3D image registration on various medical objects and imaging modalities.

Chapter 6 presents a novel 3D reconstruction method with nonlinear statistical shape model. Compared with linear SSM, nonlinear SSM can represent large shape variations between the training data. Kernel PCA is used in this work by mapping the data in the input space to a high-dimensional feature space via a nonlinear map and then applying PCA in this feature space. After linear analysis of the feature, a reverse mapping backward from feature space to input space is also non-trivial because the exact pre-image typically does not exist and only an approximation solution is available. A closed-form solution is used in this study to find the pre-image based on distance constraints in the feature space. The solution is non-iterative, involves only linear algebra, and does not suffer from local minima. To measure the similarity between different imaging modalities, a hybrid energy function is employed to combine the information from feature and intensity. The hybrid energy requires neither time-consuming DRR generation nor accurate segmentation, which is error prone.

Chapter 7 presents a knee kinematics analysis using a multi-body 2D-3D registration of the CT data to a sequence of monoplane fluoroscopic x-ray images. The

similarity measure in this study is a hybrid method that combines the information from feature, intensity, and multi-body registration term as energy function to evaluate the registration quality. For edge-image generation, prior knowledge is used to decrease impact from noises and the presence of neighboring bones.

In summary, this study proposed a novel 3D reconstruction method to reconstruct a 3D model from a sequence of monoplane fluoroscopic x-ray images with nonlinear shape prior, specifically those emphasized in the application of kinematic analysis of the knee. The proposed method can be easily extended for various medical objects and imaging modalities. Reliable 3D reconstruction opens up many exciting possibilities, including kinematics analysis, pre-operative planning, intra-operative navigation, and diagnosing purposes by comparing different data sources, such as CT, MRI and X-ray fluoroscopy images where visualizing the anatomy is fundamental.

BIBLIOGRAPHY

[1] H. Turbell and P. E. Danielsson, "Helical cone-beam tomography," *International Journal of Imaging Systems and Technology,* vol. 11, pp. 91-100, 2000.

[2] M. Mahfouz, "A New Registration Method for Implanted and Non Implanted Knee Joints Using Single Plane X-Ray Video Fluoroscopy," Ph.D., Colorado School of Mines, 2002.

[3] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "ACTIVE SHAPE MODELS - THEIR TRAINING AND APPLICATION," *Computer Vision and Image Understanding,* vol. 61, pp. 38-59, Jan 1995.

[4] R. D. K. Adrija Sharma, Mohamed R. Mahfouz, "In Vivo Kinematics Evaluation in Flexion of Patients Implanted With Primary TKA," *Techniques in Knee Surgery,* vol. 10, pp. 66-72, 2011.

[5] M. R. Mahfouz, W. A. Hoff, R. D. Komistek, and D. A. Dennis, "A robust method for registration of three-dimensional knee implant models to two-dimensional fluoroscopy images," *Ieee Transactions on Medical Imaging,* vol. 22, pp. 1561-1574, Dec 2003.

[6] N. Baka, M. de Bruijne, T. van Walsum, B. L. Kaptein, J. E. Giphart, M. Schaap, *et al.*, "Statistical Shape Model-Based Femur Kinematics From Biplane Fluoroscopy," *Ieee Transactions on Medical Imaging,* vol. 31, pp. 1573-1583, Aug 2012.

[7] N. Baka, B. L. Kaptein, M. de Bruijne, T. van Walsum, J. E. Giphart, W. J. Niessen, *et al.*, "2D-3D shape reconstruction of the distal femur from stereo X-ray imaging using statistical shape models," *Medical Image Analysis,* vol. 15, pp. 840-850, Dec 2011.

[8] S. Benameur, M. Mignotte, F. Destrempes, and J. A. De Guise, "Three-dimensional biplanar reconstruction of scoliotic rib cage using the estimation of a mixture of probabilistic prior models," *Ieee Transactions on Biomedical Engineering,* vol. 52, pp. 1713-1728, Oct 2005.

[9] M. Fleute and S. Lavallee, "Nonrigid 3-D/2-D registration of images using statistical models," in *Medical Image Computing and Computer-Assisted Intervention, Miccai'99, Proceedings*. vol. 1679, C. Taylor and A. Colchester, Eds., ed, 1999, pp. 138-147.

[10] L. Lemieux, R. Jagoe, D. R. Fish, N. D. Kitchen, and D. G. T. Thomas, "A PATIENT-TO-COMPUTED-TOMOGRAPHY IMAGE REGISTRATION METHOD BASED ON DIGITALLY RECONSTRUCTED RADIOGRAPHS," *Medical Physics,* vol. 21, pp. 1749-1760, Nov 1994.

[11] S. Benameur, M. Mignotte, S. Parent, H. Labelle, W. Skalli, and J. de Guise, "3D/2D registration and segmentation of scoliotic vertebrae using statistical models," *Computerized Medical Imaging and Graphics,* vol. 27, pp. 321-337, Sep-Oct 2003.

[12] G. Y. Zheng, "Statistical Deformable Model-Based Reconstruction of a Patient-Specific Surface Model from Single Standard X-ray Radiograph," in *Computer Analysis of Images and Patterns, Proceedings*. vol. 5702, X. Jiang and N. Petkov, Eds., ed, 2009, pp. 672-679.

[13] G. Y. Zheng, X. Dong, and M. A. G. Ballester, "Unsupervised reconstruction of a patient-specific surface model of a proximal femur from calibrated fluoroscopic

images," in *Medical Image Computing and Computer-Assisted Intervention - Miccai 2007, Pt 1, Proceedings*. vol. 4791, N. Ayache, S. Ourselin, and A. Maeder, Eds., ed, 2007, pp. 834-841.

[14] R. Bansal, L. H. Staib, Z. Chen, A. Rangarajan, J. Knisely, R. Nath, *et al.*, "A novel approach for the registration of 2D portal and 3D CT images for treatment setup verification in radiotherapy," in *Medical Image Computing and Computer-Assisted Intervention - Miccai'98*. vol. 1496, W. M. Wells, A. Colchester, and S. Delp, Eds., ed, 1998, pp. 1075-1086.

[15] P. Bifulco, M. Cesarelli, R. Allen, M. Romano, A. Fratini, and G. Pasquariello, "2D-3D Registration of CT Vertebra Volume to Fluoroscopy Projection: A Calibration Model Assessment," *Eurasip Journal on Advances in Signal Processing,* 2010.

[16] P. Bifulco, M. Cesarelli, R. Allen, M. Romano, A. Fratini, and G. Pasquariello, "2D-3D registration of CT vertebra volume to fluoroscopy projection: a calibrationmodel assessment," *EURASIP J. Adv. Signal Process,* vol. 2010, pp. 1-8, 2010.

[17] H. J. G. a. C. M. W. Devin A. Singh, "A 2D-3D Registration Tool Capable of Accurately Quantifying Flexion-Extension Positioning in the Cervical Spine," *The Open Spine Journal,* vol. 3, pp. 1-8, 2011.

[18] J. Weese, T. M. Buzug, C. Lorenz, and C. Fassnacht, "An approach to 2D/3D registration of a vertebra in 2D x-ray fluoroscopies with 3D CT images," in *Cvrmed-Mrcas'97: First Joint Conference - Computer Vision, Virtual Reality and Robotics in Medicine and Medical Robotics and Computer-Assisted Surgery*. vol. 1205, J. Troccaz, E. Grimson, and R. Mosges, Eds., ed, 1997, pp. 119-128.

[19] H. Livyatan, Z. Yaniv, and L. Joskowicz, "Gradient-based 2-D/3-D rigid registration of fluoroscopic X-ray to CT," *Ieee Transactions on Medical Imaging,* vol. 22, pp. 1395-1406, Nov 2003.

[20] P. Markelj, D. Tomazevic, F. Pernus, and B. Likar, "Robust Gradient-Based 3-D/2-D Registration of CT and MR to X-Ray Images," *Ieee Transactions on Medical Imaging,* vol. 27, pp. 1704-1714, Dec 2008.

[21] A. Szymczak, W. Hoff, and M. Mahfouz, "3D Shape from Silhouette Points in Registered 2D Images Using Conjugate Gradient Method," in *Medical Imaging 2010: Image Processing*. vol. 7623, B. M. Dawant and D. R. Haynor, Eds., ed, 2010.

[22] D. Tomazevic, B. Likar, T. Slivnik, and F. Pernus, "3-D/2-D registration of CT and MR to X-ray images," *Ieee Transactions on Medical Imaging,* vol. 22, pp. 1407-1416, Nov 2003.

[23] W. Wein, B. Roper, and N. Navab, "2D/3D registration based on volume gradients," in *Medical Imaging 2005: Image Processing, Pt 1-3*. vol. 5747, J. M. Fitzpatrick and J. M. Reinhardt, Eds., ed, 2005, pp. 144-150.

[24] J. Hermans, P. Claes, J. Bellemans, D. Vandermeulen, and P. Suetens, "Robust initialization for 2D/3D registration of knee implant models to single-plane fluoroscopy," pp. 651208-651208, 2007.

[25] M. L. Ozuysal, V. ; Fua, P. , "Pose estimation for category specific multiview object localization," *Conference on Computer Vision and Pattern Recognition*, 2009.

[26] R. Chellappa and R. Bagdazian, "FOURIER CODING OF IMAGE BOUNDARIES," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 102-105, 1984.

[27] E. Persoon and K. S. Fu, "SHAPE DISCRIMINATION USING FOURIER DESCRIPTORS," *Ieee Transactions on Systems Man and Cybernetics*, vol. 7, pp. 170-179, 1977.

[28] K. Arbter, W. E. Snyder, H. Burkhardt, and G. Hirzinger, "APPLICATION OF AFFINE-INVARIANT FOURIER DESCRIPTORS TO RECOGNITION OF 3-D OBJECTS," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 640-647, Jul 1990.

[29] C. T. Zahn and R. Z. Roskies, "FOURIER DESCRIPTORS FOR PLANE CLOSED CURVES," *Ieee Transactions on Computers*, vol. C 21, pp. 269-&, 1972.

[30] E. Berg, M. Mahfouz, C. Debrunner, and W. Hoff, "A 2D Fourier approach to deformable model segmentation of 3D medical images," in *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*. vol. 3117, M. Sonka, I. A. Kakadiaris, and J. Kybic, Eds., ed, 2004, pp. 181-192.

[31] S. Benameur, M. Mignotte, S. Parent, H. Labelle, W. Skalli, J. A. De Guise, *et al.*, "A hierarchical statistical modeling approach for the unsupervised 3D reconstruction of the scoliotic spine," in *2003 International Conference on Image Processing, Vol 1, Proceedings*, ed, 2003, pp. 561-564.

[32] 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," *Journal of Biomechanics*, vol. 36, pp. 873-882, Jun 2003.

[33] J. Boisvert, X. Pennec, H. Labelle, F. Cheriet, and N. Ayache, "Principal spine shape deformation modes using Riemannian geometry and articulated models," in *Articulated Motion and Deformable Objects, Proceedings*. vol. 4069, F. J. Perales and R. B. Fisher, Eds., ed, 2006, pp. 346-355.

[34] A. Gueziec, K. N. Wu, A. Kalvin, B. Williamson, P. Kazanzides, and R. Van Vorhis, "Providing visual information to validate 2-D to 3-D registration," *Medical Image Analysis*, vol. 4, pp. 357-374, Dec 2000.

[35] G. Zheng, "Unifying energy minimization and mutual information maximization for robust 2D/3D registration of x-ray and CT images," in *Pattern Recognition, Proceedings*. vol. 4713, F. A. Hamprecht, C. Schnorr, and B. Jahne, Eds., ed, 2007, pp. 547-557.

[36] R. Sandhu, S. Dambreville, A. Yezzi, and A. Tannenbaum, "A Nonrigid Kernel-Based Framework for 2D-3D Pose Estimation and 2D Image Segmentation," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 1098-1115, Jun 2011.

[37] O. Sadowsky, J. Lee, E. G. Sutter, S. J. Wall, J. L. Prince, and R. H. Taylor, "Hybrid Cone-Beam Tomographic Reconstruction: Incorporation of Prior

Anatomical Models to Compensate for Missing Data," *Ieee Transactions on Medical Imaging*, vol. 30, pp. 69-83, Jan 2011.

[38] M. Roth, C. Brack, R. Burgkart, A. Czopf, H. Gotte, and A. Schweikard, "Multi-view contourless registration of bone structures using a single calibrated X-ray fluoroscope," in *Cars '99: Computer Assisted Radiology and Surgery*. vol. 1191, H. U. Lemke, K. Inamura, M. W. Vannier, and A. G. Farman, Eds., ed, 1999, pp. 756-761.

[39] G. Y. Zheng, M. A. G. Ballester, M. Styner, and L. P. Nolte, "Reconstruction of patient-specific 3D bone surface from 2D calibrated fluoroscopic images and point distribution model," in *Medical Image Computing and Computer-Assisted Intervention - Miccai 2006, Pt 1*. vol. 4190, R. Larsen, M. Nielsen, and J. Sporring, Eds., ed, 2006, pp. 25-32.

[40] X. Zhang, G. Y. Zheng, F. Langlotz, and L. P. Nolte, "Assessment of spline-based 2D-3D registration for image-guided spine surgery," *Minimally Invasive Therapy & Allied Technologies*, vol. 15, pp. 193-199, Jun 2006.

[41] R. Kurazume, K. Nakamura, T. Okada, Y. Sato, N. Sugano, T. Koyama, *et al.*, "3D reconstruction of a femoral shape using a parametric model and two 2D fluoroscopic images," *Computer Vision and Image Understanding*, vol. 113, pp. 202-211, Feb 2009.

[42] H. Lamecker, T. H. Wenckebach, and H. C. Hege, "Atlas-based 3D-shape reconstruction from x-ray images," in *18th International Conference on Pattern Recognition, Vol 1, Proceedings*, Y. Y. Tang, S. P. Wang, G. Lorette, D. S. Yeung, and H. Yan, Eds., ed, 2006, pp. 371-374.

[43] D. Zikic, M. Groher, A. Khamene, and N. Navab, "Deformable registration of 3D vessel structures to a single projection image," in *Medical Imaging 2008: Image Processing, Pts 1-3*. vol. 6914, J. M. Reinhardt and J. P. W. Pluim, Eds., ed, 2008.

[44] S. Zuffi, A. Leardini, F. Catani, S. Fantozzi, and A. Cappello, "A model-based method for the reconstruction of total knee replacement kinematics," *Ieee Transactions on Medical Imaging*, vol. 18, pp. 981-991, Oct 1999.

[45] X. Chen, M. R. Varley, L.-K. Shark, G. S. Shentall, and M. C. Kirby, "An Extension of Iterative Closest Point Algorithm for 3D-2D Registration for Pre-treatment Validation in Radiotherapy," presented at the Proceedings of the International Conference on Medical Information Visualisation--BioMedical Visualisation, 2006.

[46] J. Feldmar, N. Ayache, and F. Betting, "3D-2D projective registration of free-form curves and surfaces," *Computer Vision and Image Understanding*, vol. 65, pp. 403-424, Mar 1997.

[47] L. Brunie, S. Lavallee, J. Troccaz, P. Cinquin, and M. Bolla, "PRE-IRRADIATION AND INTRA-IRRADIATION MULTIMODAL IMAGE REGISTRATION - PRINCIPLES AND FIRST EXPERIMENTS," *Radiotherapy and Oncology*, vol. 29, pp. 244-252, Nov 1993.

[48] C. Xin, M. R. Varley, L. Shark, G. S. Shentall, and M. C. Kirby, "An Extension of Iterative Closest Point Algorithm for 3D-2D Registration for Pre-treatment Validation in Radiotherapy," in *Medical Information Visualisation - BioMedical Visualisation, 2006. MediVis 2006. International Conference on*, 2006, pp. 3-8.

[49] P. J. Besl and N. D. McKay, "A METHOD FOR REGISTRATION OF 3-D SHAPES," *Ieee Transactions on Pattern Analysis and Machine Intelligence,* vol. 14, pp. 239-256, Feb 1992.

[50] T. Yamazaki, T. Watanabe, Y. Nakajima, K. Sugamoto, T. Tomita, H. Yoshikawa, *et al.*, "Improvement of depth position in 2-D/3-D registration of knee implants using single-plane fluoroscopy," *Ieee Transactions on Medical Imaging,* vol. 23, pp. 602-612, May 2004.

[51] A. Gueziec, P. Kazanzides, B. Williamson, and R. H. Taylor, "Anatomy-based registration of CT-scan and intraoperative X-ray images for guiding a surgical robot," *Ieee Transactions on Medical Imaging,* vol. 17, pp. 715-728, Oct 1998.

[52] C. R. Maurer, G. B. Aboutanos, B. M. Dawant, R. J. Maciunas, and J. M. Fitzpatrick, "Registration of 3-D images using weighted geometrical features," *Ieee Transactions on Medical Imaging,* vol. 15, pp. 836-849, Dec 1996.

[53] T. M. Buzug, J. Weese, C. Fassnacht, and C. Lorenz, "Image registration: Convex weighting functions for histogram-based similarity measures," in *Cvrmed-Mrcas'97: First Joint Conference - Computer Vision, Virtual Reality and Robotics in Medicine and Medical Robotics and Computer-Assisted Surgery*. vol. 1205, J. Troccaz, E. Grimson, and R. Mosges, Eds., ed, 1997, pp. 203-212.

[54] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *Ieee Transactions on Medical Imaging,* vol. 16, pp. 187-198, Apr 1997.

[55] C. Studholme, D. L. G. Hill, and D. J. Hawkes, *Automated 3D registration of truncated MR and CT images of the head*, 1995.

[56] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures," *Medical Physics,* vol. 24, pp. 25-35, Jan 1997.

[57] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," *International Journal of Computer Vision,* vol. 24, pp. 137-154, Sep 1997.

[58] L. M. G. Brown, T. E. Boult, and Ieee, *Registration of planar film radiographs with computed tomography*, 1996.

[59] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes, "A comparison of similarity measures for use in 2D-3D medical image registration," in *Medical Image Computing and Computer-Assisted Intervention - Miccai'98*. vol. 1496, W. M. Wells, A. Colchester, and S. Delp, Eds., ed, 1998, pp. 1153-1161.

[60] G. Y. Zheng, X. Zhang, S. Jonic, P. Thevenaz, M. Unser, and L. P. Nolte, "Point similarity measures based on MRF modeling of difference images for spline-based 2D-3D rigid registration of X-ray fluoroscopy to CT images," in *Biomedical Image Registration, Proceedings*. vol. 4057, J. P. W. Pluim, B. Likar, and F. A. Gerritsen, Eds., ed, 2006, pp. 186-194.

[61] J. Kim, S. D. Li, D. Pradhan, R. Hammoud, Q. Chen, F. F. Yin, *et al.*, "Comparison of similarity measures for rigid-body CT/Dual x-ray image registrations," *Technology in Cancer Research & Treatment,* vol. 6, pp. 337-345, Aug 2007.

[62] R. Munbodh, H. D. Tagare, Z. Chen, D. A. Jaffray, D. J. Moseley, J. P. S. Knisely, *et al.*, "2D-3D registration for prostate radiation therapy based on a statistical model of transmission images," *Medical Physics*, vol. 36, pp. 4555-4568, Oct 2009.

[63] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes, "A comparison of similarity measures for use in 2-D-3-D medical image registration," *Ieee Transactions on Medical Imaging*, vol. 17, pp. 586-595, Aug 1998.

[64] R. H. Gong, A. J. Stewart, and P. Abolmaesumi, "A new method for CT to fluoroscope registration based on unscented Kalman filter," in *Medical Image Computing and Computer-Assisted Intervention - Miccai 2006, Pt 1*. vol. 4190, R. Larsen, M. Nielsen, and J. Sporring, Eds., ed, 2006, pp. 891-898.

[65] T. Rohlfing, D. Russakoff, J. Denzler, and C. Maurer, Jr., "Progressive Attenuation Fields: Fast 2D-3D Image Registration Without Precomputation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004*. vol. 3216, C. Barillot, D. Haynor, and P. Hellier, Eds., ed: Springer Berlin Heidelberg, 2004, pp. 631-638.

[66] W. Birkfellner, R. Seemann, M. Figl, J. Hummel, C. Ede, P. Homolka, *et al.*, "Wobbled splatting - a fast perspective volume rendering method for simulation of x-ray images from CT," *Physics in Medicine and Biology*, vol. 50, pp. N73-N84, May 2005.

[67] W. Birkfellner, M. Stock, M. Figl, C. Gendrin, J. Hummel, S. Dong, *et al.*, "Stochastic rank correlation: A robust merit function for 2D/3D registration of image data obtained at different energies," *Medical Physics*, vol. 36, pp. 3420-3428, Aug 2009.

[68] J. Weese, R. Gocke, G. P. Penney, P. Desmedt, T. M. Buzug, and H. Schumann, "Fast voxel-based 2D/3D registration algorithm using a volume rendering method based on the shear-warp factorization," in *Medical Imaging 1999: Image Processing, Pts 1 and 2*. vol. 3661, K. M. Hanson, Ed., ed, 1999, pp. 802-810.

[69] D. Knaan and L. Joskowicz, "Effective Intensity-Based 2D/3D Rigid Registration between Fluoroscopic X-Ray and CT," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003*. vol. 2878, R. Ellis and T. Peters, Eds., ed: Springer Berlin Heidelberg, 2003, pp. 351-358.

[70] X. L. Li, J. Yang, and Y. M. Zhu, "Digitally reconstructed radiograph generation by an adaptive Monte Carlo method," *Physics in Medicine and Biology*, vol. 51, pp. 2745-2752, Jun 2006.

[71] LaRose, "Iterative x-ray/CT registration using accelerated volume rendering," Th.D., Robotics Institute, Carnegie Mellon University, 2001.

[72] T. S. Y. Tang, N. J. MacIntyre, H. S. Gill, R. A. Fellows, N. A. Hill, D. R. Wilson, *et al.*, "Hardware-assisted 2D/3D intensity-based registration for assessing patellar tracking," in *Medical Image Computing and Computer-Assisted Intervention - Miccai 2004, Pt 2, Proceedings*. vol. 3217, C. Barillot, D. R. Haynor, and P. Hellier, Eds., ed, 2004, pp. 1095-1096.

[73] S. M. Hofstetter R, Sati M, Nolte LP., "Fluoroscopy as an imaging means for computer-assisted surgical navigation.," *Computer Aided Surgery,* vol. 4, pp. 65-76, 1999.

[74] C. L. Baker, "COMPUTED TOMOGRAPHY FROM IMAGERY GENERATED BY FLUOROSCOPY ALONG AN ARBITRARY PATH," Master, Colorado School of Mines, 2004.

[75] S. A. Banks and W. A. Hodge, "Accurate measurement of three-dimensional knee replacement kinematics using single-plane fluoroscopy," *Ieee Transactions on Biomedical Engineering,* vol. 43, pp. 638-649, Jun 1996.

[76] E. E. A. Fatah, "Three Dimensional Nonlinear Statistical Modeling Framework for Morphological Analysis," doctoral, The University of Tennessee, 2012.

[77] J. B. Stiehl, R. D. Komistek, D. A. Dennis, R. D. Paxson, and W. A. Hoff, "FLUOROSCOPIC ANALYSIS OF KINEMATICS AFTER POSTERIOR-CRUCIATE-RETAINING KNEE ARTHROPLASTY," *Journal of Bone and Joint Surgery-British Volume,* vol. 77B, pp. 884-889, Nov 1995.

[78] T. E. McAlindon, C. Cooper, J. R. Kirwan, and P. A. Dieppe, "KNEE PAIN AND DISABILITY IN THE COMMUNITY," *British Journal of Rheumatology,* vol. 31, pp. 189-192, Mar 1992.

[79] G. Behiels, D. Vandermeulen, F. Maes, P. Suetens, and P. Dewaele, "Active Shape Model-based segmentation of digital X-ray images," in *Medical Image Computing and Computer-Assisted Intervention, Miccai'99, Proceedings*. vol. 1679, C. Taylor and A. Colchester, Eds., ed, 1999, pp. 128-137.

[80] C. Lindner, S. Thiagarajah, J. M. Wilkinson, G. A. Wallis, and T. F. Cootes, "Fully Automatic Segmentation of the Proximal Femur Using Random Forest Regression Voting," *Ieee Transactions on Medical Imaging,* vol. 32, pp. 1462-1472, Aug 2013.

[81] J. Mu, X. M. Liu, S. Luan, P. H. Heintz, G. W. Mlady, and D. Z. Chen, "Segmentation of Knee Joints in X-ray Images Using Decomposition-based Sweeping and Graph Search," *Medical Imaging 2011: Image Processing,* vol. 7962, 2011.

[82] S. J. M. Matthias Seise, Ian W. Ricketts and Carlos A. Wigderowitz, "segmenting tibia and femur from knee x-ray images," *Medical Image Understanding and Analysis*, pp. 103-106, 2005.

[83] P. C. Mahalanobis, "On the generalised distance in statistics," *Proceedings of the National Institute of Sciences of India,* vol. 2, pp. 49–55, 1936.

[84] J. B. Shi and J. Malik, "Normalized cuts and image segmentation," *Ieee Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, pp. 888-905, Aug 2000.

[85] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *Ieee Transactions on Pattern Analysis and Machine Intelligence,* vol. 23, pp. 681-685, Jun 2001.

[86] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour Detection and Hierarchical Image Segmentation," *Ieee Transactions on Pattern Analysis and Machine Intelligence,* vol. 33, pp. 898-916, May 2011.

[87] C. Fowlkes, D. Martin, and J. Malik, "Learning affinity functions for image segmentation: combining patch-based and gradient-based approaches," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, 2003, pp. II-54-61 vol.2.

[88] G. H. G. a. C. F. V. Loan, *Matrix Computations*: John Hopkins Press, 1989.

[89] S. Belongie and J. Malik, "Finding boundaries in natural images: A new method using point descriptors and area completion," in *Computer Vision — ECCV'98*. vol. 1406, H. Burkhardt and B. Neumann, Eds., ed: Springer Berlin Heidelberg, 1998, pp. 751-766.

[90] B. C. M. a. M. R. Mahfouz, "Unsupervised three-dimensional segmentation of medical images using an anatomical bone atlas," presented at the Int. Conf. Biomed. Engr, Singapore, 2005.

[91] P. J. Besl and H. D. McKay, "A method for registration of 3-D shapes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 14, pp. 239-256, 1992.

[92] F. L. Bookstein, "Landmark methods for forms without landmarks: localizing group differences in outline shape," *Medical Image Analysis,* vol. 1, pp. 225-244, 1997.

[93] T. F. C. a. C. J. Taylor., "Statistical Models of Appearance for Computer Vision.," in *Tech. Report*, ed. University of Manchester,, 2000.

[94] L. Najman and M. Schmitt, "Geodesic saliency of watershed contours and hierarchical segmentation," *Ieee Transactions on Pattern Analysis and Machine Intelligence,* vol. 18, pp. 1163-1173, Dec 1996.

[95] A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures " *Analytical Chemistry,* vol. 36, pp. 1627-&, 1964.

[96] L. R. Dice, "MEASURES OF THE AMOUNT OF ECOLOGIC ASSOCIATION BETWEEN SPECIES," *Ecology,* vol. 26, pp. 297-302, 1945.

[97] G. Li, S. K. Van de Velde, and J. T. Bingham, "Validation of a non-invasive fluoroscopic imaging technique for the measurement of dynamic knee joint motion," *J Biomech,* vol. 41, pp. 1616-22, 2008.

[98] Z. L. Zhu and G. A. Li, "An automatic 2D-3D image matching method for reproducing spatial knee joint positions using single or dual fluoroscopic images," *Computer Methods in Biomechanics and Biomedical Engineering,* vol. 15, pp. 1245-1256, 2012.

[99] S. Miao, J. Lucas, and R. Liao, "Automatic Pose Initialization for Accurate 2-D/3-D Registration Applied to Abdominal Aortic Aneurysm Endovascular Repair," in *Medical Imaging 2012: Image-Guided Procedures, Robotic Interventions, and Modeling*. vol. 8316, D. R. Holmes and K. H. Wong, Eds., ed, 2012.

[100] E. K. Antonsson and R. W. Mann, "AUTOMATIC 6-DOF KINEMATIC TRAJECTORY ACQUISITION AND ANALYSIS," *Journal of Dynamic Systems Measurement and Control-Transactions of the Asme,* vol. 111, pp. 31-39, Mar 1989.

[101] M. Sati, J. A. de Guise, S. Larouche, and G. Drouin, "Quantitative assessment of skin-bone movement at the knee," *The Knee,* vol. 3, pp. 121-138, 1996.

[102] 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*, vol. 38, pp. 241-253, Feb 2005.

[103] E. B. van de Kraats, G. P. Penney, D. Tomazevic, T. van Walsum, and W. J. Niessen, "Standardized evaluation methodology for 2-D-3-D registration," *Ieee Transactions on Medical Imaging*, vol. 24, pp. 1177-1189, Sep 2005.

[104] A. Thayananthan, "Template-based Pose Estimation and Tracking of 3D Hand Motion," Doctor of Philosophy, University of Cambridge, 2005.

[105] A. Khotanzad and J. J. H. Liou, "Recognition and pose estimation of unoccluded three-dimensional objects from a two-dimensional perspective view by banks of neural networks," *Ieee Transactions on Neural Networks*, vol. 7, pp. 897-906, Jul 1996.

[106] W. A. Hoff, R. D. Komistek, D. A. Dennis, S. M. Gabriel, and S. A. Walker, "Three-dimensional determination of femoral-tibial contact positions under in vivo conditions using fluoroscopy," *Clinical Biomechanics*, vol. 13, pp. 455-472, Oct 1998.

[107] C.-C. C. a. C.-j. Llin, "LIBSVM : a library for support vector machine," 2001.

[108] M. A. Aizerman, Braverma.Em, and L. I. Rozonoer, "THEORETICAL FOUNDATIONS OF POTENTIAL FUNCTION METHOD IN PATTERN RECOGNITION LEARNING," *Automation and Remote Control*, vol. 25, pp. 821-&, 1965.

[109] V. V. C. Cortes, "SUPPORT-VECTOR NETWORKS," *Machine Learning*, vol. 20, pp. 273-297, 1995.

[110] E. E. Abdel Fatah, N. R. Shirley, R. L. Jantz, and M. R. Mahfouz, "Improving sex estimation from crania using a novel three-dimensional quantitative method," *J Forensic Sci*, vol. 59, pp. 590-600, May 2014.

[111] M. R. Mahfouz, B. C. Merkl, E. E. Fatah, R. Booth, Jr., and J. N. Argenson, "Automatic methods for characterization of sexual dimorphism of adult femora: distal femur," *Comput Methods Biomech Biomed Engin*, vol. 10, pp. 447-56, Dec 2007.

[112] M. Mahfouz, A. Badawi, B. Merkl, E. E. A. Fatah, E. Pritchard, K. Kesler, *et al.*, "Patella sex determination by 3D statistical shape models and nonlinear classifiers," *Forensic Science International*, vol. 173, pp. 161-170, Dec 2007.

[113] F. P. Kuhl and C. R. Giardina, "Elliptic Fourier features of a closed contour," *Computer Graphics and Image Processing*, vol. 18, pp. 236-258, 1982.

[114] N. S. Altman, "AN INTRODUCTION TO KERNEL AND NEAREST-NEIGHBOR NONPARAMETRIC REGRESSION," *American Statistician*, vol. 46, pp. 175-185, Aug 1992.

[115] B. Scholkopf, Smola, A., *Learning with kernels*. Cambridge, MA: MIT Press, 2002.

[116] D. Glasner, M. Galun, S. Alpert, R. Basri, and G. Shakhnarovich, "Viewpoint-aware object detection and continuous pose estimation," *Image and Vision Computing*, vol. 30, pp. 923-933, Dec 2012.

[117] C. Y. Zhao, H. X. Zhang, X. Y. Zhang, R. S. Zhang, F. Luan, M. C. Liu, *et al.*, "Prediction of milk/plasma drug concentration (MIP) ratio using support vector machine (SVM) method," *Pharmaceutical Research,* vol. 23, pp. 41-48, Jan 2006.

[118] Z. Fei and T. Tiyip, "A Method of Soil Salinization Information Extraction with SVM Classification Based on ICA and Texture Features," *Agricultural Science & Technology*, pp. 1046-1049+1074, 2011.

[119] S. a. P. Knerr, et al, "Single-Layer Learning Revisited: A Stepwise. Procedure for Building and Training a Neural Network," in *Neurocomputing: Algorithms, Architectures and Applications*. vol. F68, F. Souli, Ed., ed: Springer-Verlag, 1990, pp. 41-50.

[120] S. C. Chen and R. F. Murphy, "A graphical model approach to automated classification of protein subcellular location patterns in multi-cell images," *Bmc Bioinformatics,* vol. 7, Feb 2006.

[121] K. V. M. I.L.Dryden. (1998). *Statistical Shape Analysis*.

[122] J. M. Fitzpatrick, J. B. West, and C. R. Maurer, "Predicting error in rigid-body point-based registration," *Ieee Transactions on Medical Imaging,* vol. 17, pp. 694-702, Oct 1998.

[123] M. J. van der Bom, L. W. Bartels, M. J. Gounis, R. Homan, J. Timmer, M. A. Viergever, *et al.*, "Robust initialization of 2D-3D image registration using the projection-slice theorem and phase correlation," *Medical Physics,* vol. 37, pp. 1884-1892, Apr 2010.

[124] W. Y. Qi, L. X. Gu, Q. Zhao, and Ieee, "Effective 2D-3D Medical Image Registration using Support Vector Machine," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vols 1-8*, ed, 2008, pp. 5386-5389.

[125] C. Li, A. Hosseini, T.-Y. Tsai, Y.-M. Kwon, and G. Li, "Articular contact kinematics of the knee before and after a cruciate retaining total knee arthroplasty," *Journal of orthopaedic research : official publication of the Orthopaedic Research Society,* vol. 33, pp. 349-58, 2015 Mar (Epub 2014 Dec 2015.

[126] R. D. Komistek, D. A. Dennis, and M. Mahfouz, "In vivo fluoroscopic analysis of the normal human knee," *Clinical Orthopaedics and Related Research,* pp. 69-81, May 2003.

[127] 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,* vol. 31, pp. 10-16, Jan 2009.

[128] M. R. Torry, K. B. Shelburne, D. S. Peterson, J. E. Giphart, J. P. Krong, C. Myers, *et al.*, "Knee Kinematic Profiles during Drop Landings: A Biplane Fluoroscopy Study," *Medicine and Science in Sports and Exercise,* vol. 43, pp. 533-541, Mar 2011.

[129] E. S. Abebe, G. M. Utturkar, D. C. Taylor, C. E. Spritzer, J. P. Kim, C. T. Moorman, III, *et al.*, "The effects of femoral graft placement on in vivo knee kinematics after anterior cruciate ligament reconstruction," *Journal of Biomechanics,* vol. 44, pp. 924-929, Mar 15 2011.

[130] S. Mika, B. Scholkopf, A. Smola, K. R. Muller, M. Scholz, and G. Ratsch, "Kernel PCA and de-noising in feature spaces," in *Advances in Neural Information Processing Systems 11*. vol. 11, M. S. Kearns, S. A. Solla, and D. A. Cohn, Eds., ed, 1999, pp. 536-542.

[131] J. T. Y. Kwok and I. W. H. Tsang, "The pre-image problem in kernel methods," *Ieee Transactions on Neural Networks,* vol. 15, pp. 1517-1525, Nov 2004.

[132] H. V. Franc V, "Statistical pattern recognition toolbox for MATLAB," Czech Technical University Prague2004.

[133] W. Yu, "The Convergence Property of the Simplex Evolutionary Techniques," *Scientia Sinica,* vol. Special Issue of Mathematics, 1, pp. 68-77, // 1979.

[134] T. Netsch, P. Rosch, A. van Muiswinkel, and J. Weese, "Towards real-time multi-modality 3-D medical image registration," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 2001, pp. 718-725 vol.1.

[135] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis,* vol. 2, pp. 1-36, 3// 1998.

[136] S. Dambreville, Y. Rathi, and A. Tannenbaum, "A framework for image segmentation using shape models and kernel space shape priors," *Ieee Transactions on Pattern Analysis and Machine Intelligence,* vol. 30, pp. 1385-1399, Aug 2008.

[137] J. Wu, E. E. A. Fatah, and M. R. Mahfouz, "Fully automatic initialization of two-dimensional–three-dimensional medical image registration using hybrid classifier," *Journal of Medical Imaging,* vol. 2, pp. 024007-024007, 2015.

[138] B. J. Fregly, W. G. Sawyer, M. K. Harman, and S. A. Banks, "Computational wear prediction of a total knee replacement from in vivo kinematics," *Journal of Biomechanics,* vol. 38, pp. 305-314, Feb 2005.

VITA

Jing Wu was born in Shanghai, China in 1983. She attended Shanghai Jiaotong University in 2001, receiving her Bachelor of Science and Master of Science degree in Material Science and Engineering in 2005 and 2008, respectively, while performing computer vision and control theory research applied to welding robotics under Dr. Shanben Chen at the Intellgentized Robotic Welding Technology Laboratory. She then came to United States for her Ph.D. at the University of Tennessee under Dr. Mohamed Mahfouz at the Center for Musculoskeletal Research. Her Ph.D. work focused on computer vision and machine learning applied to medical imaging.

Professionally, Jing's research interests remain focused on applications of medical imaging pertaining to issues related to the human body, specifically in computer vision, signal processing, machine learning, and optimization.