Fully automatic initialization of two-dimensional–three-dimensional medical image registration using hybrid classifier

Jing Wu
Emam E. Abdel Fatah
Mohamed R. Mahfouz

Fully automatic initialization of two-dimensional–three-dimensional medical image registration using hybrid classifier

Jing Wu, Emam E. Abdel Fatah, and Mohamed R. Mahfouz*
University of Tennessee, Institute of Biomedical Engineering, 1506 Middle Drive, Knoxville, Tennessee 37996-2000, United States

Abstract. X-ray video fluoroscopy along with two-dimensional–three-dimensional (2D-3D) registration techniques is widely used to study joints *in vivo* kinematic behaviors. These techniques, however, are generally very sensitive to the initial alignment of the 3-D model. We present an automatic initialization method for 2D-3D registration of medical images. The contour of the knee bone or implant was first automatically extracted from a 2-D x-ray image. Shape descriptors were calculated by normalized elliptical Fourier descriptors to represent the contour shape. The optimal pose was then determined by a hybrid classifier combining *k*-nearest neighbors and support vector machine. The feasibility of the method was first validated on computer synthesized images, with 100% successful estimation for the femur and tibia implants, 92% for the femur and 95% for the tibia. The method was further validated on fluoroscopic x-ray images with all the poses of the testing cases successfully estimated. Finally, the method was evaluated as an initialization of a feature-based 2D-3D registration. The initialized and uninitialized registrations had success rates of 100% and 50%, respectively. The proposed method can be easily utilized for 2D-3D image registration on various medical objects and imaging modalities. © *2015 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JMI.2.2.024007]

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.[1–3] Figure 1 shows the 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 1-min protocol gives the patient a radiation exposure on the order of 0.6 to 1.8 "rad equivalent man" (rem).[3] However, two-dimensional (2-D) fluoroscopic images do not provide three-dimensional (3-D) information. A 3-D CAD/computed tomography (CT)/MRI model is time-consuming (MRI) to generate and exposes patients to more radiation (CT), but it provides 3-D insight and improves visualization of anatomical structures. Therefore, it is necessary to fuse the information of the 2-D images and 3-D volume data by 2D-3D registration.

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.[2] Different groups have investigated the use of dual fluoroscopy as a method for increasing the out-of-plane registration accuracy.[4] However, limited studies have investigated the differences between the two methods using a consistent 2D-3D matching setup.[5]

The key step in 2D-3D registration is to determine the relationship between the 2-D x-ray image coordinate system and the 3-D model coordinate system. This can be achieved by manual registration of 3-D data to be fitted to the 2-D 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.[6–8] However, skin markers are susceptible to error due to undesired motion between the 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 so as to maximize a similarity measure, which reflects the quality of the registration. A large body of work of 2D-3D registration based on 3-D model registration to radiographs[2,9–11] and 3-D kinematic analysis from fluoroscopic images[12] has shown satisfactory registration accuracy. Existing 2D-3D registration methods are often limited by a small capture range.[13] Due to the nonconvex nature of the similarity measure, such methods can suffer from inaccurate registration results when the initial pose of the 3-D 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 a 2-D image with a precomputed training model to determine

*Address all correspondence to: Mohamed R. Mahfouz, E-mail: mmahfouz@ibme.utk.edu

Fig. 1 Fluoroscopic imaging in the knee.

the initial pose of the 3-D 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 registrations.

There is a wide variety of published literature on the topic of pose estimation based on template matching.[14–17] Khotanzad and Liou[15] developed a neural network-based system for recognition and pose estimation of an unoccluded 3-D object from any single 2-D 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 a global and unique solution.[18]

Banks and Hodge[16] 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 et al.[17] 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 3-D 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 dataset. Instead, we used a hybrid classifier to determine the pose of the 3-D model so as to avoid a blanket search. More specifically, the hybrid classifier combines k-nearest neighbors (k-NN) and SVM, where k-NN serves as an attention mechanism, generating an area of interest for the further local search by SVM.[19,20] 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 subproblems. 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 rest of the paper is organized as follows. The approaches are described in Sec. 2. Section 3 provides the experiments and their results. Finally, Sec. 4 concludes the paper and gives future prospects.

2 Approach

2.1 Problem Statement

The goal of the proposed method is to estimate the poses of the 3-D model from a single plane fluoroscopic x-ray image based on a precomputed training set of 2-D contours. This method helps kinematics in that it can determine 3-D *in vivo* and weight bearing kinematics of the knee joint from a 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 3-D 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 3-D 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

$$X = T(X_0) = RX_0 + t, \qquad (1)$$

where $R \in SO(3)$ is a rotation matrix $R = R_\gamma \times R_\beta \times R_\alpha$,

$$R_\alpha = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}, \qquad R_\beta = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix},$$

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

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

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

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.[21] ASM represents the global shape constraints with the dominant shape variation in the training set. The benefit of ASM is that it constrains the search to feasible shapes 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 3-D surface mesh models projection onto the 2-D plane at different poses, yielding 1887 training contours. The contours are then aligned by finding corresponding points of each contour using statistical atlases, where training contours
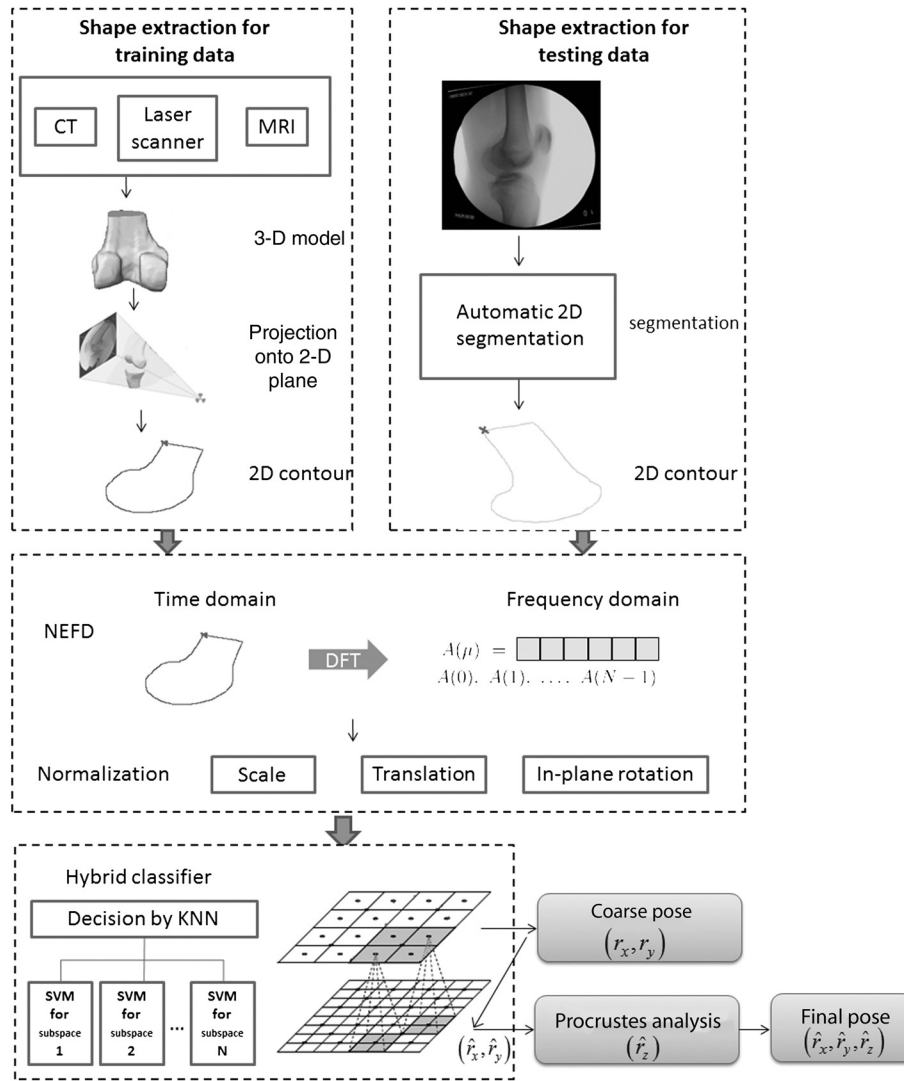
Fig. 2 Flow chart of the initialization system.

with a known point distribution are deformed to match one another.[22–24] To segment the x-ray image, we start from the average contour in the training dataset; search on the normal profile of each contour point for the greatest gradient; use ASM to constrain the contour to a 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 Fig. 3.

An example of segmentation results is shown in Fig. 4.

2.3 Shape Representation

The shape representation is based on NEFDs developed by Kuhl and Giardina.[25] 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.

2.3.1 Calculate shape descriptors of the contour

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

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

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

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

$$d_n = \frac{T}{2n^2\pi^2} \sum_{i=1}^{k} \frac{dY_i}{dt_i} \left[\sin\frac{2n\pi t_i}{T} - \sin\frac{2n\pi t_{i-1}}{T} \right], \quad (5)$$
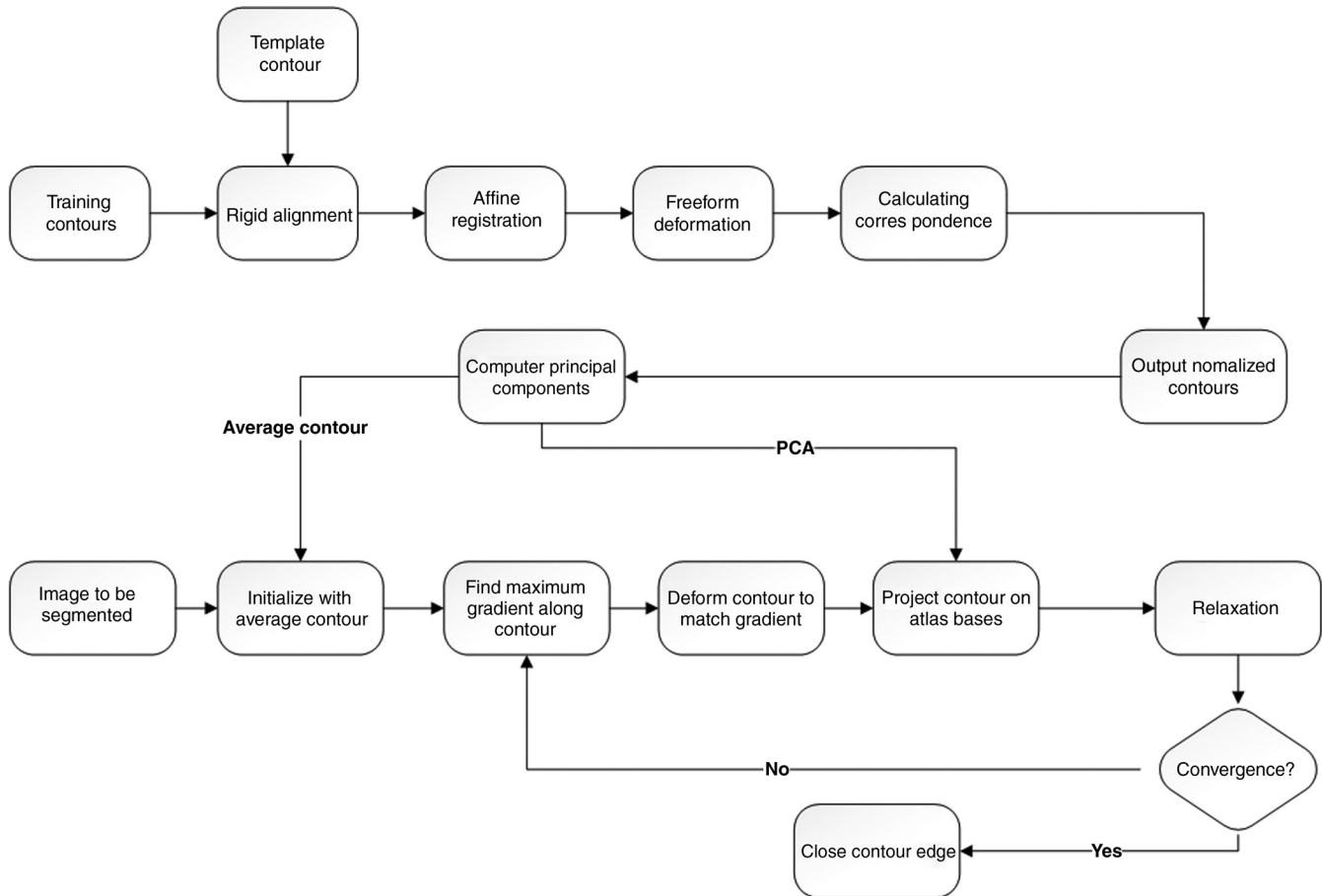
Fig. 3 Flow chart of the two-dimensional (2-D) segmentation.

Fig. 4 Segmentation result of femur and tibia in an x-ray image.

where X_i and Y_i are the X and Y coordinate values; a_i and b_i represent the projection on the semi major and semi minor x axis, and c_i and d_i similarly represent 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, 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

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

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

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 Fig. 5(a), where root-mean-square (RMS) error between the reconstructed and the original contour of the femur drops as the number of harmonics increases. Forty harmonics were found to provide accurate representation for the training shapes according to the reconstruction errors shown in Fig. 5(a). The comparison
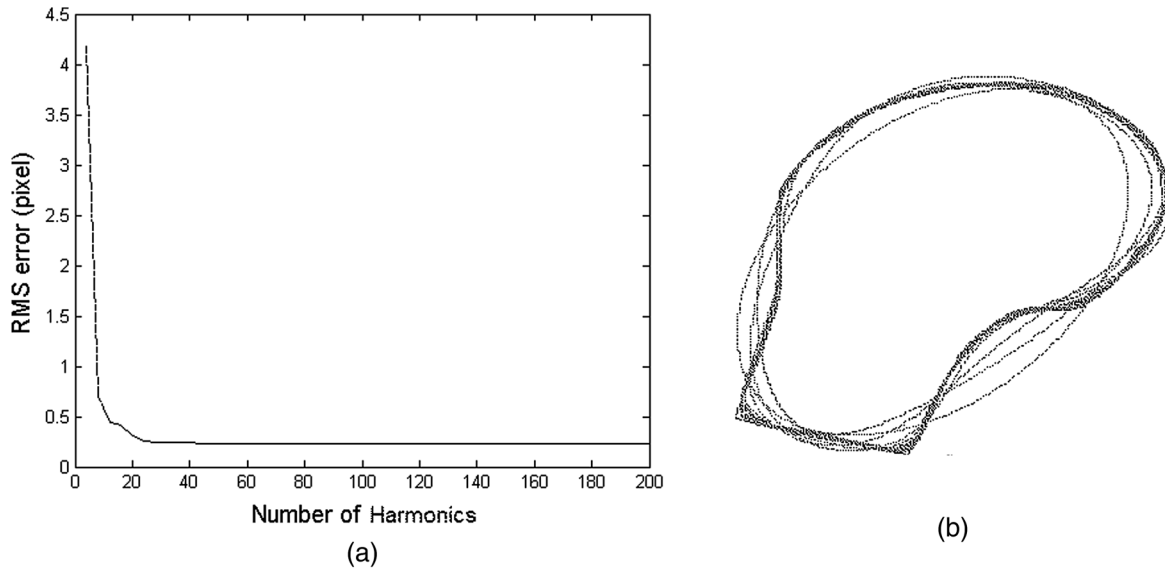
Fig. 5 Reconstruction error of NEFD: (a) root-mean-square (RMS) error of reconstructed femur contours with respect to the number of harmonics, (b) reconstructed femur contours are in red with the number of harmonics from 1 to 50. The original contour is in blue.

between the original and reconstructed contours from Fourier descriptors is shown in Fig. 5(b).

2.3.2 Normalization

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

$$X_i' = X_i + \frac{1}{k}\sum_{n=1}^{k} X_n, \tag{8}$$

$$Y_i' = Y_i + \frac{1}{k}\sum_{n=1}^{k} Y_n. \tag{9}$$

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

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

where $\theta_1 = (1/2)\mathrm{arctg}[2(a_1 b_1 + c_1 d_1)/a_1^2 + c_1^2 - b_1^2 - d_1^2]$ 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:

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

where $\phi_1 = \mathrm{arctg}(c_1^*/a_1^*)$.

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

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

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

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-NN and SVM classifiers. k-NN is a nonparametric 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-NNs.[26] While k-NN is natural in this kind of problem, it suffers from high variance in the case of limited sampling. Moreover, since the NN 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 the classes. But SVM involves time-consuming optimization and computation of pairwise distances, which makes it difficult to classify a large number of classes. We thus combine k-NN and SVM to improve the overall performance.

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

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

where $\mathrm{Ed}(A, B)$ is the Euclidean distance between instance descriptors A and B, and s_{Ai} and s_{Bi} are the i'th NEFD of instances 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 a maximal margin.[27] When it is not possible to linearly separate the two classes, kernel function is used to map the data into a high dimensional space where the data are linearly separable. Given training vectors $x_i \in R^n$, $i = 1, \ldots, l$, 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.

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

(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 radial-basis function (RBF) kernel (or Gaussian kernel) is the most commonly used kernel. Glasner et al. reported a good performance of the RBF kernel[28] in the pose estimation problem. We used a Gaussian RBF kernel, $\Phi(x_i \cdot x_j) = \exp(-\gamma \|x_i - x_j\|^2)$.

Penalty parameter C in Eq. (14) and γ in the 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 vectors. 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.[29] The technique used to determine its optimal parameters is a grid search using a cross-validation[18,30] 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 multiclasses include one-against-all method, one against one,[31] and directed acyclic graph.[32] We use the one-against-one method[31] in this paper.

Scaling of the feature vectors is also one of the important factors affecting performance.[18] The training set was normalized by scaling it into [0, 1] and then test data were normalized with the same degree of scaling.

2.5 Estimate In-Plane Rotation

Equation (11) in Sec. 2.3.2 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.[33]

$$\theta_{z\,\text{est}} = \theta_{z\,\text{input}} - \theta_\Delta.$$

(15)

Before Procrustes analysis, corresponding points across contours must be established using statistical atlases, as with the corresponding point procedures discussed in Sec. 2.2.

3 Experiments and Results

Three sets of experiments were performed to validate the proposed method. In the first set of experiments (Sec. 3.2), computer simulated images were used to estimate the 3DOF transformation parameters (rotation in x, y, and z axes). A comparison was made among the accuracy of three different classifiers. In the second set of experiments (Sec. 3.3), we further tested the initialization method with x-ray images. In the third set of experiments (Sec. 3.4), the initialization of 6DOF transformation parameters (translation in x, y, and z axes; rotation in x, y, and z axes) was followed by a feature-based 2D-3D registration algorithm. A comparison of performance between initialized and uninitialized registrations was made to test whether the proposed method leads to improved registration.

3.1 Training Data

Two datasets were used for training and testing. The first is journey cruciate substituting (BCS) prostheses (Smith and Nephew, Memphis, Tennessee)[34] composed of 33 femoral and tibial implants. The implants were laser scanned using a Digibot III machine (DIGI-BOTICS Austin, Texas), which recorded the vertices on the implants' surface. The second dataset is William M. Bass Donated Skeletal Collection of 100 femurs and tibias from the University of Tennessee's Anthropology Department. 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 were 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 3-D models is depicted in Fig. 6.
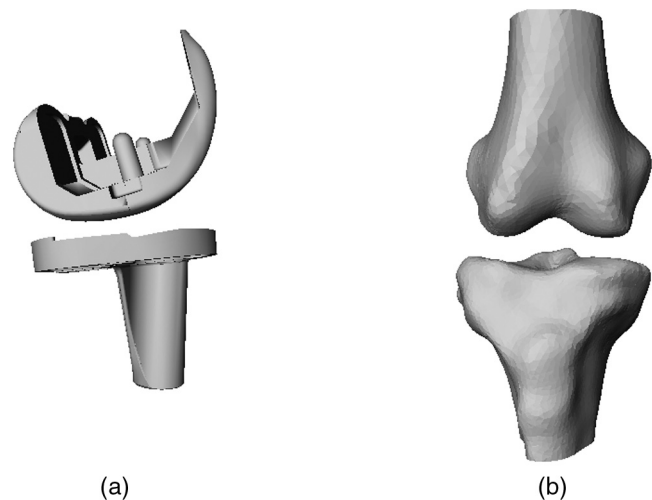

(a)

(b)

Fig. 6 Three-dimensional (3-D) models of knee bone and implant: (a) 3-D knee implant; (b) 3-D knee bone.
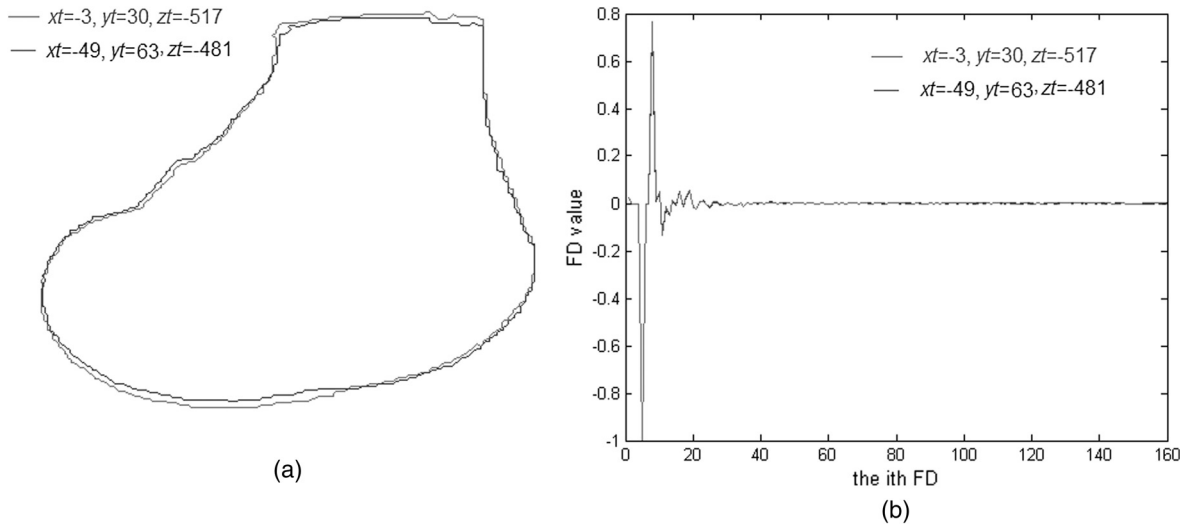
Fig. 7 Perspective effect by comparing two poses with the greatest variation in translation and fixed rotation during deep knee bending (DKB): (a) projected shapes; (b) NEFD of the projected shapes.

To generate the training set, the 3-D models were rendered on the 2-D plane by a software developed by the authors[2] using a 3-D graphic library (Open Inventor, TGS, San Diego, California). A coordinate system was established with the x axis pointing to the left, y axis pointing upward, 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 poses of the DKB, respectively. As can be seen from Figs. 7(a) and 7(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.

The 3-D models were rotated in 1-deg increments about the x and y axes while fixing the position and in-plane rotation. The rotational range is $x \in [190,170]$, $y \in [-20,20]$ for the femoral implant; $x \in [-15,5]$, $y \in [-15,20]$ for the tibial implant; $x \in [-80,-40]$, $y \in [-175,-150]$ for the femur; and $x \in [80,110]$, $y \in [-30,0]$ for the tibia. At each orientation, the 3-D models were rendered as a binary image so that only the silhouettes are visible.

3.2 Experiments on Simulated Data

Computer simulated data of femoral implant, tibial implant, femur, and tibia were generated by rendering the 3-D models onto the 2-D 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 the x and y axes and varying in the range of a 40 deg difference along the z axis. The increment was 1 deg for each of the three axial rotations. 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[13,35] Since a rotation of 1 deg around one of

the three axes will lead to an mTRE of 0.8–1.0 mm,[36] estimates with an angular error smaller than 4 deg can be safely considered successful. The capture range was defined as the 95% success range as was proposed in Ref. 13.

Ten-fold cross-validations were performed to test the performance of the initialization method. For femoral and tibial implants, 28 patients' data were used for training and five patients' data for testing. For femur and tibia, 90 patients' data were used for training and 10 patients' data for testing. Each patient's data contains an equal number of poses as class labels. Table 1 presents the estimates by the initialization method for femoral implant, tibial implant, femur, and tibia. By setting the success threshold as the 4-deg angular difference between estimated and gold standard poses,[13,35] 100% of the femoral and tibial implants; 92% of the femur and 95% of the tibia estimation were successful.

A comparison of performance among single k-NN, single SVM, and k-NN-SVM is presented in Table 2. The performance improves significantly from single k-NN to k-NN-SVM. k-NN-SVM also outperforms single SVM. Moreover, k-NN-SVM is much faster to train because each SVM only involves samples in the local neighborhood, and the number of classes for k-NN-SVM is much smaller than that of single SVM.

Table 1 Root-mean-square (RMS) error of synthetic data experiments with k-NN-SVM classifier using 10-fold cross validation (90% patients for training and 10% patients for testing).

	Xr (deg)	Yr (deg)	Zr (deg)	Success rate (%)
Femoral implant	1.24 ± 0.99	0.75 ± 0.59	1.13 ± 1.12	100
Tibia implant	0.73 ± 0.66	0.44 ± 0.45	1.85 ± 1.86	100
Femur	3.02 ± 1.76	2.32 ± 1.83	1.66 ± 1.16	92
Tibia	1.45 ± 1.55	2.74 ± 1.94	1.53 ± 1.23	95

Table 2 Average error of tibia pose estimation by *k*-NN, SVM, and *k*-NN-SVM classifier (90% patients for training and 10% patients for testing).

	Xr (deg)	Yr (deg)	Zr (deg)	Success rate (%)
k-NN-SVM	1.45 ± 1.55	2.74 ± 1.94	1.53 ± 1.23	95
k-NN	2.90 ± 1.75	4.24 ± 2.06	2.09 ± 1.22	81
SVM	1.71 ± 1.58	3.06 ± 2.17	1.54 ± 1.24	95

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. 8.

Contours were automatically extracted using the segmentation algorithm mentioned in Sec. 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 DKB were used for validation from both datasets. The average error of femoral implant, tibial implant, femur, and tibia is
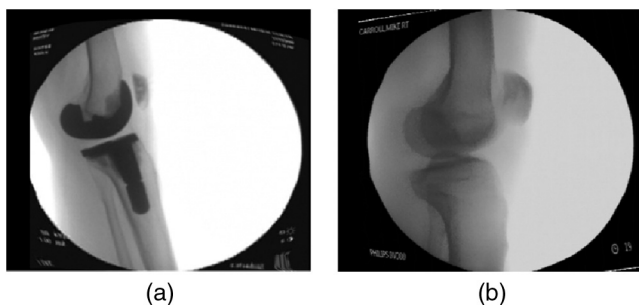

(a)   (b)

Fig. 8 Fluoroscopic x-ray images: (a) femoral and tibial implant; (b) femur and tibia.

Table 3 Average error of experiments on x-ray images. Twenty x-ray images for each case.

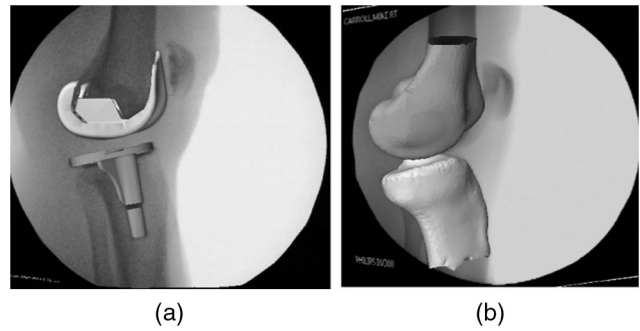|	Xr(deg)	Yr(deg)	Zr(deg)
Femoral implant	1.10 ± 0.57	0.70 ± 0.82	1.10 ± 0.31
Tibial implant	0.70 ± 0.51	0.42 ± 0.53	1.75 ± 0.42
Femur	0.60 ± 1.00	1.3 ± 1.49	0.80 ± 0.63
Tibia	1.50 ± 0.52	1.90 ± 0.32	1 ± 1.25

(a)   (b)

Fig. 9 Overlay of the estimated 3-D model on the 2-D x-ray image: (a) femoral and tibial implant, (b) femur and tibia.

shown in Table 3, where all the prediction poses are within the tolerance angle of 4 deg. An overlay of the estimated 3-D model on the 2-D x-ray image is shown in Fig. 9.

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 registrations. The feature-based 2D-3D registration algorithm was implemented with a normalized cross-correlation similarity measure[2] and optimization was performed by simulated annealing. Registrations with a final error smaller than 1 deg 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 the *x* and *y* axes was estimated by the center of the contour. Translation in the *z* axis was estimated by comparing the scale ratio of the testing contour with the one of the known pose.[32]

A comparison between the result of uninitialized and initialized registrations was made using 10 cases of femoral implant, tibia implant, femur, and tibia, as shown in Table 4. 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.

3.5 Speed of the System

The time required for pose estimation was about 5 s on average and the time required for training SVM models was about 15 h. 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,[37] whose registration time is 375.3 s and training time is 473.2 s, our method has an obvious advantage in the pose estimation speed. The training of SVM models with our method requires a longer time,[37] but it is an offline process thus does not impact the performance of the online pose estimation. The training speed can be improved by parallelizing the training process in the future.

4 Discussions and Conclusions

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 2-D contour from *x*-ray images. Normalized elliptical Fourier descriptors were used

Table 4 Results of initialized and uninitialized registrations.

		Average error (deg)	Average error (deg) (obvious failure removed)	Correct registration (%)	Initialization registration	
					Average error (deg)	Correct registration (%)
Uninitialized registration (starting point at the middle of the angle range)						
Femur implant	Xr	5.03 ± 7.07	0.83 ± 0.61	30	0.00 ± 0.00	100
	Yr	9.72 ± 12.08	0.76 ± 0.85		0.00 ± 0.00	
	Zr	1.94 ± 2.92	0.83 ± 1.31		0.01 ± 0.03	
Tibia implant	Xr	2.67 ± 4.91	0.5 ± 0.74	40	0.14 ± 0.13	100
	Yr	0.98 ± 1.43	0.89 ± 1.4		0.12 ± 0.13	
	Zr	1.96 ± 4.01	0.24 ± 0.34		0.01 ± 0.06	
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 implant	Xr	25.36 ± 1.32	1.60 ± 1.98	10		
	Yr	29.98 ± 23.54	0.45 ± 0.35			
	Zr	14.29 ± 12.25	0.85 ± 0.78			
Tibia implant	Xr	19.55 ± 38.01	1.40 ± 1.98	10		
	Yr	13.81 ± 14.65	1.65 ± 1.91			
	Zr	18.81 ± 24.29	0.85 ± 1.20			
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			

to represent the shape of the 2-D contour. Then a hybrid classifier integrating k-NN and SVM estimated the pose of the 3-D 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.[5] 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 registrations. Reliable 2D-3D registration opens up many exciting possibilities such as preoperative planning, intraoperative 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 deg in each direction). Therefore, this method is only suitable for initialization purposes. However, it can search in a large range of possible solutions, leading to a capture range as large as the domain of the training data. Moreover, the proposed initialization method does not need any information from the 3-D model during testing. Another limitation is related to mirror and circular symmetry, which leads to similar 2-D projection images at symmetric views. Currently, the proposed method does not 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.

Acknowledgments

References

1. J. B. Stiehl et al., "Fluoroscopic analysis of kinematics after posterior-cruciate-retaining knee arthroplasty," *J. Bone Joint Surg. Br.* **77**(6), 884–889 (1995).
2. M. R. Mahfouz et al., "A robust method for registration of three-dimensional knee implant models to two-dimensional fluoroscopy images," *IEEE Trans. Med. Imaging.* **22**, 1561–1574 (2003).
3. R. D. K. Adrija Sharma and M. R. Mahfouz, "In vivo kinematics evaluation in flexion of patients implanted with primary TKA," *Tech. Knee Surg.* **10**(2), 66–72 (2011).
4. 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.* **41**, 1616–1622 (2008).
5. 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," *Comput. Methods Biomech. Biomed. Eng.* **15**, 1245–1256 (2012).
6. 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*, D. R. Holmes and K. H. Wong, Eds., Vol. 8316, (2012).
7. E. K. Antonsson and R. W. Mann, "Automatic 6-DOF kinematic trajectory acquisition and analysis," *J. Dyn. Syst. Meas. Control* **111**, 31–39 (1989).
8. M. Sati et al., "Quantitative assessment of skin-bone movement at the knee," *Knee* **3**, 121–138, (1996).
9. N. Baka et al., "Statistical shape model-based femur kinematics from biplane fluoroscopy," *IEEE Trans. Med. Imaging* **31**, 1573–1583 (2012).
10. G. Zheng, "Unifying energy minimization and mutual information maximization for robust 2D/3D registration of x-ray and CT images," in *Proc. Pattern Recognition*, vol. 4713, F. A. Hamprecht, C. Schnorr, and B. Jahne, Eds., pp. 547–557 (2007).
11. P. Bifulco et al., "2D-3D registration of CT vertebra volume to fluoroscopy projection: a calibration model assessment," *EURASIP J. Adv. Signal Process.* **2010**, 1–8 (2010).
12. D. A. Dennis et al., "In vivo determination of normal and anterior cruciate ligament-deficient knee kinematics," *J. Biomech.* **38**, 241–253 (2005).
13. E. B. van de Kraats et al., "Standardized evaluation methodology for 2-D-3-D registration," *IEEE Trans. Med. Imaging* **24**, 1177–1189 (2005).
14. A. Thayananthan, "Template-based pose estimation and tracking of 3D hand motion," Doctor of Philosophy, University of Cambridge (2005).
15. 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 Trans. Neural Netw.* **7**, 897–906 (1996).
16. S. A. Banks and W. A. Hodge, "Accurate measurement of three-dimensional knee replacement kinematics using single-plane fluoroscopy," *IEEE Trans. Biomed. Eng.* **43**, 638–649 (1996).
17. W. A. Hoff et al., "Three-dimensional determination of femoral-tibial contact positions under in vivo conditions using fluoroscopy," *Clin. Biomech.* **13**, 455–472 (1998).
18. C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machine," *Acm Trans. Intel. Syst. Technol.* **2**(3), 1–27 (2001).
19. M. A. Aizerman, E. M. Braverma, and L. I. Rozonoer, "Theoretical foundations of potential function method in pattern recognition learning," *Autom. Remote Control* **25**(6), 821–837 (1964).
20. V. V. C. Cortes, "Support-vector networks," *Mach. Learn.* **20**(3), 273–297 (1995).
21. T. F. Cootes et al., "Active shape models: their training and application," *Comput. Vis. Image Und.* **61**, 38–59 (1995).
22. E. E. Abdel Fatah et al., "Improving sex estimation from crania using a novel three-dimensional quantitative method," *J. Forensic Sci.* **59**, 590–600 (2014).
23. M. R. Mahfouz et al., "Automatic methods for characterization of sexual dimorphism of adult femora: distal femur," *Comput. Methods Biomech. Biomed. Eng.* **10**, 447–456 (2007).
24. M. Mahfouz et al., "Patella sex determination by 3D statistical shape models and nonlinear classifiers," *Forensic Sci. Int.* **173**, 161–170 (2007).
25. F. P. Kuhl and C. R. Giardina, "Elliptic Fourier features of a closed contour," *Comput. Graph. Image Process.* **18**, 236–258 (1982).
26. N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Am. Stat.* **46**(3), 175–185 (1992).
27. B. Scholkopf and A. Smola, *Learning with Kernels*, MIT Press, Cambridge, Massachusetts (2002).
28. D. Glasner et al., "Viewpoint-aware object detection and continuous pose estimation," *Image Vis. Comput.* **30**, 923–933 (2012).
29. C. Y. Zhao et al., "Prediction of milk/plasma drug concentration (MIP) ratio using support vector machine (SVM) method," *Pharm. Res.* **23**, 41–48 (2006).
30. Z. Fei and T. Tiyip, "A method of soil salinization information extraction with SVM classification based on ICA and texture features," *Agri. Sci. Technol.* **12**(7), 1046–1049+1074 (2011).
31. S. Knerr, L. Personnaz, and G. Dreyfus, "Single-layer learning revisited: a stepwise procedure for building and training a neural network," in *Neurocomputing: Algorithms, Architectures and Applications*, F. Souli, Ed., Vol. F68, Springer-Verlag, pp. 41–50 (1990).
32. S. C. Chen and R. F. Murphy, "A graphical model approach to automated classification of protein subcellular location patterns in multicell images," *Bmc Bioinf.* **7**(1), 90 (2006).
33. I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*, Wiley, Chichester (1998).
34. J. Victor et al., "In vivo kinematics after a cruciate-substituting TKA," *Clini. Orthop. Relat. Res.* **468**, 807–814 (2010).
35. J. M. Fitzpatrick, J. B. West, and C. R. Maurer, "Predicting error in rigid-body point-based registration," *IEEE Trans. Med. Imaging* **17**, 694–702 (1998).
36. M. J. van der Bom et al., "Robust initialization of 2D-3D image registration using the projection-slice theorem and phase correlation," *Med. Phys.* **37**, 1884–1892 (2010).
37. W. Y. Qi et al., "Effective 2D-3D medical image registration using support vector machine," in *2008 30th Annual Int. Conf. Proc. IEEE Eng. Med. Biol. Soc.*, Vol. 1–8, pp. 5386–5389 (2008).

Jing Wu received her bachelor's and master's degrees in material science and engineering (MSE) from Shanghai Jiao Tong University in 2005 and 2008. Currently, she is working toward her PhD in biomedical engineering at the University of Tennessee. Her current research interests include medical imaging, computer vision, signal processing, machine learning, and optimization.

Emam E. Abdel Fatah received his bachelor's degree in biomedical engineering from Cairo University and his PhD in biomedical engineering from the University of Tennessee. Currently, he is an assistant professor at the Institute of Biomedical Engineering iBME at the University of Tennessee. He has published various articles in distinguished international journals, with over 50 abstracts and short papers accepted for publication at international conferences. His research interests include medical imaging, machine learning, biomechanics, anthropology, and pervasive computing.

Mohamed R. Mahfouz is a professor in the Mechanical, Aerospace, and Biomedical Engineering Departments at the University of Tennessee (UT). He was awarded the prestigious titles of research fellow by UT's College of Engineering and chair, career development professor in 2007 and 2011. Currently, he is chair of the Institute of Biomedical Engineering at UT. He has published 150+ journal articles in distinguished journals, with 400+ abstracts and short papers, and has given 40+ guest lectureships worldwide.