FACTS: Fully Automatic CT Segmentation of a Hip Joint

CHENGWEN CHU, CHENG CHEN, LI LIU, and GUOYAN ZHENG

Institute for Surgical Technology and Biomechanics, University of Bern, Stauffacherstr. 78, 3014 Bern, Switzerland

Associate Editor Sean S. Kohles oversaw the review of this article.

Abstract—Extraction of surface models of a hip joint from CT data is a pre-requisite step for computer assisted diagnosis and planning (CADP) of periacetabular osteotomy (PAO). Most of existing CADP systems are based on manual segmentation, which is time-consuming and hard to achieve reproducible results. In this paper, we present a Fully Automatic CT Segmentation (FACTS) approach to simultaneously extract both pelvic and femoral models. Our approach works by combining fast random forest (RF) regression based landmark detection, multi-atlas based segmentation, with articulated statistical shape model (aSSM) based fitting. The two fundamental contributions of our approach are: (1) an improved fast Gaussian transform (IFGT) is used within the RF regression framework for a fast and accurate landmark detection, which then allows for a fully automatic initialization of the multi-atlas based segmentation; and (2) aSSM based fitting is used to preserve hip joint structure and to avoid penetration between the pelvic and femoral models. Taking manual segmentation as the ground truth, we evaluated the present approach on 30 hip CT images (60 hips) with a 6-fold cross validation. When the present approach was compared to manual segmentation, a mean segmentation accuracy of 0.40, 0.36, and 0.36 mm was found for the pelvis, the left proximal femur, and the right proximal femur, respectively. When the models derived from both segmentations were used to compute the PAO diagnosis parameters, a difference of $2.0 \pm 1.5°$, $2.1 \pm 1.6°$, and $3.5 \pm 2.3\%$ were found for anteversion, inclination, and acetabular coverage, respectively. The achieved accuracy is regarded as clinically accurate enough for our target applications.

Keywords—Segmentation, Random forest regression, Gaussian transform, Atlas, Periacetabular osteotomy, Acetabular coverage, Hip.

INTRODUCTION

Developmental dysplasia of the hip (DDH) is a congenital defect that has serious effects on patients.

DDH is characterized by mal-formed hip acetabulum (socket) which is about 20% smaller than a normal case, and therefore patients with DDH exhibit insufficient acetabular coverage (AC) and high joint contact pressure.[17] Here the AC is defined as a ratio between the femoral head surface covered by the acetabulum and the complete surface. Poor AC and high joint contact pressure consequently lead to complications such as degenerative osteoarthritis which affects patients with limited range of motion and great pain.[17]

Among many DDH treatment options, periacetabular osteotomy (PAO) has become a common surgical intervention.[8] Aiming to increase AC and reduce joint contact pressure, PAO is implemented by reorienting the acetabulum to realign it with the femoral head. In PAO surgery, knowing the required rotation for realignment of the acetabulum is very important since poor reorientation will fail the surgery with unimproved AC. However, in clinical routine it is a challenge for surgeons to decide the optimal rotation before surgery. Therefore, computer aided diagnosis and planning (CADP) systems for PAO are expected to address this challenge.[16] With such a CADP system, operation can be simulated by virtually rotating acetabulum and analyzing the AC in different orientations. Thus, it is possible to find the optimal rotation angle. One of the basic and critical functions of such a system is to extract surface models of both pelvis and proximal femur from hip CT data. Since manual segmentation is time-consuming and hard to achieve reproducible results, it is the goal of this study to develop a fully automatic hip CT segmentation algorithm.

The existing work on hip CT segmentation can be largely classified into three categories: (a) intensity based methods,[4,10,13,29] (b) statistical shape model (SSM) based methods,[11,12,14,20,26,27] and (c) atlas based methods.[7,19] Due to the relatively higher intensity values of bone tissues when compared to other

Address correspondence to Guoyan Zheng, Institute for Surgical Technology and Biomechanics, University of Bern, Stauffacherstr. 78, 3014 Bern, Switzerland. Electronic mail: guoyan.zheng@ieee.org

surrounding tissues, there exist works focusing on bone segmentation using basic image processing techniques.[4,10,13,29] These methods suffer from the difficulty in handling gray level inhomogeneities and diffused boundaries around the joint region, leading to insufficient segmentation accuracy.

To address this challenge, both SSM based methods[11,12,14,20,26,27] and atlas based methods [7,19] have been introduced. The segmentation within these methods is usually done with a two-step procedure. Firstly, the target structures have to be detected in the volume data and secondly, surface models of the target structures need to be segmented. Existing work on hip CT segmentation addresses the first step either by assuming a user-supplied initialization with manually specified anatomical landmarks [26,27] or by using generalized Hough transform (GHT).[12,20] After that, the segmentation can be performed automatically. SSM based methods start with construction of a SSM of the target structure from a given training population. The constructed SSM is then subsequently fitted to a new image to get the segmentation. Different from SSM based methods, atlas based segmentation methods employ prior knowledge about shape and intensity distributions by registering a set of atlases to a given volume. Here we define an atlas as a pair of data consisting of a CT volume and its corresponding manual segmentation. Both single atlas and multiple atlases based methods have been proposed before. Although the multi-atlases based methods have been shown to achieve improved robustness and accuracy over the single-atlas based method in brain tissue segmentation,[1] abdominal organ segmentations,[5,23] and SSM creation,[24] only single atlas based methods have been used for hip CT segmentation.[7,19]

Most of existing methods on hip CT segmentation focus on segmenting a single anatomical structure without considering its relationship to the neighboring structures.[12,14,20,26,27] When those techniques would be applied to segmentation of patient-specific models of a hip joint, the segmented models may penetrate each other due to narrowness of the hip joint space and hence do not represent a true hip joint of the patient. This has motivated the introduction of the so-called articulated statistical shape model (aSSM).[11] By approximately modeling the hip joint as a ball-and-socket joint, joint posture can be modeled as a parameterized rotation of the femur around the joint center. Fitting such an aSSM to a given hip CT volume will have the advantage of preserving hip joint structure.

In this paper, we present a fully automatic hip CT segmentation framework combing random forest (RF) regression based landmark detection, multi-atlas based segmentation, with aSSM based fitting. RF regression

based methods[3] have been proved to be efficient in landmark detection from medical image[6,15] and are also known to be computationally intensive when applied to 3D data. Here we propose to use an improved fast Gaussian transform (IFGT)[25] within the RF regression framework for a fast and accurate landmark detection. The detected landmarks then allow for not only a robust initialization of the atlases within the target volume space but also an effective selection of a subset of atlases for a fast and accurate atlas based segmentation. Joint structure of the segmented models is preserved by fitting an aSSM of the hip joint to the multi-atlas based segmentation results.

MATERIALS AND METHODS

Figure 1 shows an overview of the complete workflow of our approach. It starts with the landmark detection based on RF regression. The detected landmarks allow one to conduct paired-point registrations to align all atlases with the target image. After the registration, a subset of atlases with the least paired-point registration errors are then selected and used for a multi-atlas based segmentation. Finally, the hip joint surface models are extracted by fitting an aSSM to the multi-atlas based segmentation results.

RF Regression Based Landmark Detection

RF regression based landmark detection works in two stages: training stage and prediction stage. In the training stage, we train a regressor using a set of labeled training data. In the prediction stage, a landmark is detected by estimating the displacements from a set of randomly sampled local image patches to the unknown position of the landmark based on patch appearance, and the individual predictions are then combined in a voting scheme to predict the landmark position.

(A) Basic Algorithm: We have a separate RF landmark detector for each landmark. During training, in each training image, we sample a set of image subvolumes around the known landmark position. Each sampled sub-volumes is represented by its visual feature $\mathbf{f}_i \in \mathbb{R}^{d_f}$ and the displacement $\mathbf{d}_i \in \mathbb{R}^3$ from its center to the landmark (Fig. 2a). Let us denote training data as $\{P_i = (\mathbf{f}_i, \mathbf{d}_i)\}_{i=1...N}$ (Fig. 2b). The goal is then to learn a mapping function $\phi : \mathbb{R}^{d_f} \to \mathbb{R}^3$ from the feature space to the displacement space. Principally, any regression method can be used. In this paper, we utilize the random forest regressors.[3]

Once the regressor is trained, given a new image (Fig. 2c) in the prediction stage, we randomly sample another set of sub-volumes $\{P'_k = (\mathbf{f}'_k, \mathbf{c}'_k)\}_{i=1...N'}$ all
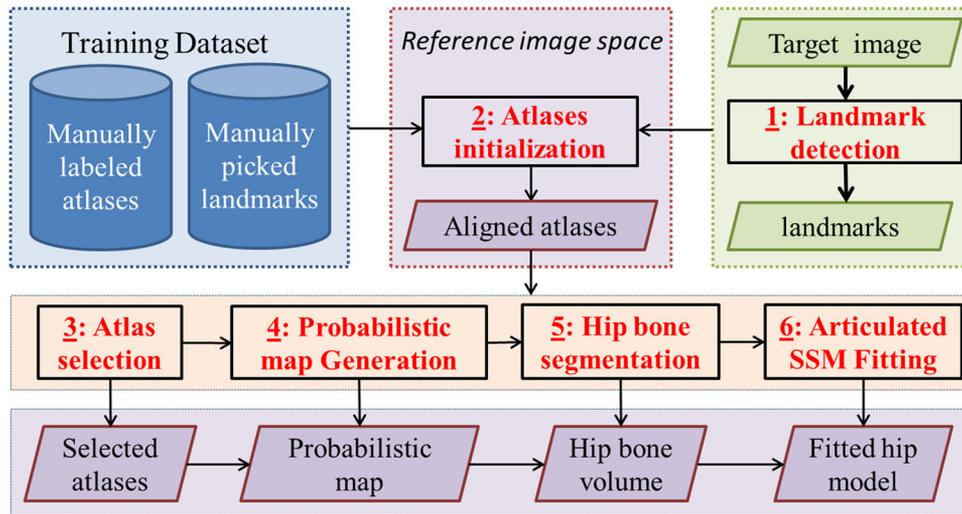
FIGURE 1.   The flowchart of the present fully automatic segmentation approach.
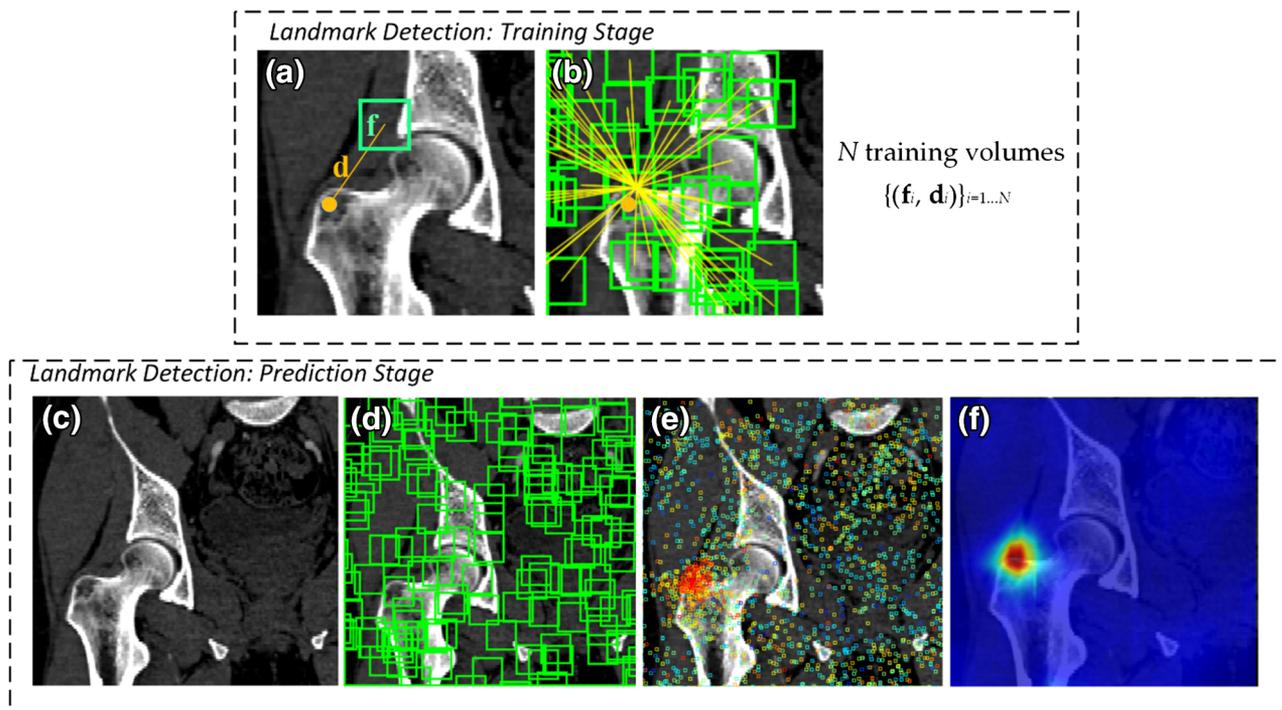

FIGURE 2.   A schematic illustration of the RF training and regression for landmark detection (illustrated on a 2D coronal slice for an easy understanding but all computations are done in 3D). (a) A sub-volume sampled around the known landmark position in a training atlas. (b) Multiple sampled sub-volumes from a training atlas. (c) A target image. (d) Sampled sub-volumes over the target image. (e) Each sub-volume gives a single vote. (f) Response image calculated from all votes.

over the image (or a region of interest (ROI) if an initial guess of the landmark position is known), where $\mathbf{f}'_k$ and $\mathbf{c}'_k$ are the visual feature and the center coordinate of the k-th sub-volume, respectively (Fig. 2d). Using the trained regressor ϕ, we can calculate displacement $\mathbf{d}'_k = \phi(\mathbf{f}'_k)$ and then $\mathbf{d}'_k + \mathbf{c}'_k$ becomes the prediction of the landmark position by a single volume P'_k (Fig. 2e). Note that each tree in the random forest will return a prediction. Therefore, supposing that there are t trees in the forest, we will get $N' \times t$ predictions. These individual predictions are very noisy, but when combined, they approach an accurate prediction. To this end, we consider each single vote as a small Gaussian distribution. We developed a fast probability aggregating algorithm as described below to sum these distributions to get a soft probability map

called *response image* which gives, for every position of the image, its probability of being the landmark (Fig. 2f).

(B) Fast Probability Aggregating: As described above, $N' \times t$ predictions are produced to detect each landmark. We consider each prediction a Gaussian model $\mathcal{N} \sim (\bar{\mathbf{d}}'_k, \sum(\mathbf{d}'_k))$ and $\sum(\mathbf{d}'_k) = diag(\sigma^2_{k,x}\sigma^2_{k,y}\sigma^2_{k,z})$ are mean and covariance (which can be calculated from the displacements of the training samples that arrived at particular leaf node). All the $N' \times t$ predictions are accumulated to compute the response image for each landmark. Once the response image has been obtained for each landmark, the position mode is selected as the landmark position.

The computation time of landmark detection depends on multivariate Gaussian accumulation which is usually computed using

$$G(\mathbf{y}_i) = \sum_{k=1}^{N' \times t} \frac{1}{\sqrt{(2\pi)^3 |\sum \mathbf{d}'_k|}} \exp(-\frac{1}{2}(\mathbf{d}_{\mathbf{y}_i} - \bar{\mathbf{d}}'_k)^T \quad (1)$$
$$\sum(\mathbf{d}'_k)^{-1}(\mathbf{d}_{\mathbf{y}_i} - \bar{\mathbf{d}}'_k)$$

where $\mathbf{d}_{\mathbf{y}_i} = \mathbf{y}_i - \mathbf{c}'_k$, $\mathbf{y}_i$ is a voxel in target image and $\mathbf{c}'_k$ is the center of volume k. For all of the N_l landmarks, such a calculation on a 3D CT image with M voxels will result in prohibitively expensive computation time of $O(M \times N' \times t \times N_l)$. In this paper, we propose to approximate Eq. (1) by:

$$G(\mathbf{y}_i) = \sum_{k}^{N' \times t} W_k \cdot e^{(\|\mathbf{d}_{\mathbf{y}_i} - \bar{\mathbf{d}}'_k\|^2/h^2)} \quad (2)$$

Here we rewrite the Eq. (1) by introducing a constant window size of h, and we move the variances out of the exponential part by introducing a weight $W_k = 1/(\sigma_{x,k}\sigma_{y,k}\sigma_{z,k})$. With such an approximation, we develop a fast strategy based on the improved fast Gaussian transform (IFGT)[25] to calculate the response images with highly reduced computational time of $O((M + N' \times t) \times N_l)$.

(C) Fast Visual Feature Computing: To calculate the visual feature for the sampled sub-volumes, we first divide each sampled sub-volume using a grid of $q \times q \times q$ blocks (Fig. 3). For each block within this sampled sub-volume, we use mean and variance of intensity as the feature, leading to a set of features with dimension of $2q^3$ for each sampled sub-volume.

To accelerate the feature extraction within each block, we use the well-known integrate image technique.[22] Details about how to compute the integral image of a quantity can be found in.[22] The quantity can be the voxel intensity value or any arithmetic computation on the intensity value. Advantage of using integral image lies in the fact that once we obtain an integral image of the quantity over the complete hip CT volume, the sum of the quantity in any sub-volume can be calculated quickly in constant time $O(1)$ no matter how big the size of the volume is.[22] Here we assume that we already computed the integral images of the voxel intensity I and the integral images of the squared voxel intensity S of the complete hip CT volume using the technique introduced in.[22] We then compute the mean $E[X]$ and the variance $Var(X)$ of the intensity value of any block (Fig. 3, right) as:

$$E[X] = (I(\mathbf{h}) - I(\mathbf{d}) - I(\mathbf{f}) - I(\mathbf{g}) + I(\mathbf{b})$$
$$+ I(\mathbf{c}) + I(\mathbf{e}) - I(\mathbf{a}))/N$$

$$E[X^2] = (S(\mathbf{h}) - S(\mathbf{d}) - S(\mathbf{f}) - S(\mathbf{g}) + S(\mathbf{b})$$
$$+ S(\mathbf{c}) + S(\mathbf{e}) - S(\mathbf{a}))/N \quad (3)$$

$$Var(X) = E[X^2] - (E[X])^2$$

where $\{\mathbf{a}, \ldots \mathbf{h}\} \in \mathcal{R}^3$ are the eight vertices of the block and N is the number of voxels within the block.

Atlas Initialization and Selection

In this step, using the detected N_l anatomical landmarks, paired-point scaled rigid registrations are per-
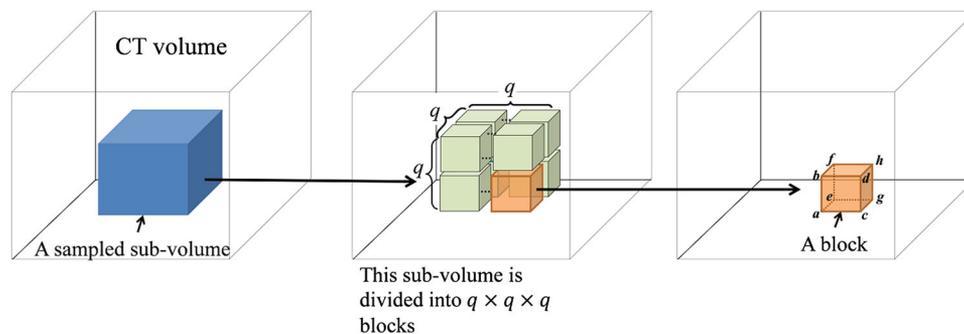

FIGURE 3. A schematic illustration on how to compute the visual feature of a sampled sub-volume for RF training and regression. Left: a sub-volume is sampled from a hip joint CT volume. Middle: we divide the sampled sub-volume into $q \times q \times q$ blocks. Right: for each block, we compute its mean and variance using the integral image technique.

formed to align all the N_A atlases to the target image space. Here, we speeded up the scaled rigid registrations by aligning all the atlases and the target image to the same reference image, which is randomly selected as one of the training images. Since all the atlases have already been aligned to the reference image prior to the segmentation phase, we only need to transform target image to the reference space. Based on the scaled rigid registration results, we select N_S atlases with the least paired-point registration errors for the given target image. The selected N_S atlases are then registered to the target image using a discrete optimization based non-rigid registration.[9]

Multi-Atlas Based Segmentation

In this step, we first use the selected N_S atlases after registration to generate probabilistic atlas (PA) both for background and hip bone tissues as shown in Fig. 4. We then formulate the multi-atlas based segmentation problem as a Maximum A Posteriori (MAP) estimation which can be efficiently solved with a graph cut based optimization method.[2] Using the known label of the selected atlases, PA of the target image is computed as:

$$\Lambda_p(l) = \frac{\sum_{i=0}^{N_s} \omega(A_i, \mathcal{T}) \delta(L_i^p, l)}{\sum_{i=0}^{N_s} \omega(A_i, \mathcal{T})} \qquad (4)$$

where

$$\delta(l, l') = \begin{cases} 1 & if \ l = l' \\ 0 & otherwise \end{cases} \qquad (5)$$

Here, i represents the index of selected atlases, $\mathbf{p}$ represents the voxel in CT image, l represents the label of bones, $L_i^{\mathbf{p}}$ is the label of the voxel $\mathbf{p}$ in the atlas A_i and

$\Lambda_{\mathbf{p}}(l)$ is probability that voxel p labeling as l. $\omega(A_i, \mathcal{T})$ is the weight of atlas A_i which evaluated by the image similarity between the atlas and the target image. We use the normalized cross correlation (NCC) as the atlas-target similarity. The probability $\Lambda_{\mathbf{p}}(l)$ is calculated for every voxels in the target image for background (where $l = 0$), pelvis (where $l = 1$), left femur (where $l = 2$), and right femur (where $l = 3$), respectively.

For the given target image, a voxel-wise MAP estimation is defined as

$$L_{\mathbf{p}} = arg \max_l P(l|I_{\mathbf{p}}) \qquad (6)$$

where $L_{\mathbf{p}}$ is the given label of the voxel $\mathbf{p}$, $I_{\mathbf{p}}$ is the intensity of the $\mathbf{p}$, l is the label of each, and $P(l|I_{\mathbf{p}})$ is the posterior probability. The MAP estimation aims to find a label l which can maximize the posterior probability. In other words, for a voxel $\mathbf{p}$ which has intensity $I_{\mathbf{p}}$, if the label l of any region can maximize the posterior probability, the voxel will be assigned a label of l. The posterior probability is computed according the Bayes theory as $P(l|I_{\mathbf{p}}) \propto p(I_{\mathbf{p}}|l)\Lambda_{\mathbf{p}}(l)$, where $\Lambda_{\mathbf{p}}(l)$ is the PA describing the prior probability and $p(I_{\mathbf{p}}|l)$ is the conditional probability that is computed as we did in.[5] The MAP estimation is then solved with the graph cut method.[2] For details, we refer to.[2]

Hip Joint Reconstruction Using a SSM Based Fitting

(A) SSM Definition: In order to build a SSM we assume a set of n aligned training shapes s_i, where $i \in \{0, \ldots, n-1\}$. Each shape s_i is described by a vector containing m vertices: $s_i = \{x_0, y_0, z_0, x_1, y_1, z_1, \ldots, x_{m-1}, y_{m-1}, z_{m-1}\}$.. By performing Principal Component Analysis (PCA) on the training set, the shape variations can be described by a linear model:
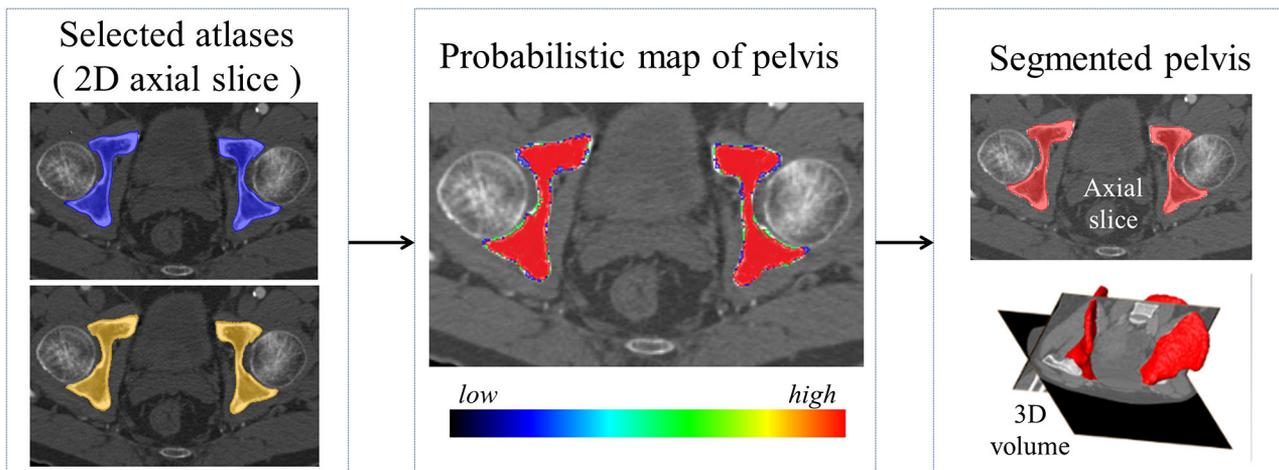

FIGURE 4. Example of generating PA for segmentation. Left: after non-rigidly registered two selected atlases to the target space. Middle: generated PA. Right: segmented pelvis.

$$S(\mathbf{b}) = \bar{s} + \sum_{i=0}^{n-2} b_i p_i \qquad (7)$$

where $\bar{s}$ is the mean shape vector, p_i the eigenvectors spanning the principal directions of the shape space and $b = (b_0, b_1, \ldots, b_{n-2})$ the shape coefficient vector.

(B) aSSM Definition: A compound model was built containing two objects P and F and a rotation center c. Whereas object P and F represents the pelvis- and femur-model respectively. We denote a vertex in a shape instance $S(\mathbf{b})$ as $v_j(\mathbf{b}), j \in \{0, 1, \ldots, m-1\}$. $v_j(\mathbf{b})$ with $j \in \{0, 1, \ldots, u-1\}$ belong to object P, whereas u denotes the number of vertices of P. Object F's vertices are $v_j(\mathbf{b}), j \in \{u, u_{+1}, \ldots, m-2\}$, and the joint center is defined as $v_c(\mathbf{b})$ and $c = m - 1$. Following,[11] we approximate the relation between the femur and the pelvis as a ball-and-socket joint, the joint posture is explicitly parameterized by a rotation R of object F around the joint center $v_c(\mathbf{b})$ which defines the relative transformation of the femur to the pelvis. Furthermore, assuming a scaled rigid transformation T between the shape space and the input image space, our parameterized aSSM is described as

$$S(\mathbf{b}, T, R) = (T \circ R(v_c(\mathbf{b})))(S(\mathbf{b})) \qquad (11)$$

where $R(v_c(\mathbf{b}))$ describes the rotation of all vertices on the proximal femur around the joint center $v_c(\mathbf{b})$ and will be only applied to those vertices $v_j(\mathbf{b}), j \in \{u, u_{+1}, \ldots, m-2\}$ on the instantiated femoral object and T is the scaled rigid transformation that will be applied to all vertices on the instantiated compound model.

(C) aSSM Construction: The correspondences between the training shapes were established using a templating method. To obtain the same vertex ordering for each bone the training shapes were reconstructed from a single template mesh by displacing its vertices to other shapes in the training set with a nonrigid transformation. To build the template mesh, one of the CT volumes was selected as the initial reference. The non-rigid transformation was calculated using the diffeomorphic demons algorithm.[21]

(D) aSSM Based Model Instantiation: Given the atlas based segmentation and the aSSM, we first extract edge points from the binary segmentation and then instantiate a compound model from the aSSM and the extract edge points using the model instantiation method introduced in[11] but with a Mahalanobis distance-based regularization as we did in.[28] Two adaptations, i.e., the transformation adaptation where the scaled rigid transformation T and the rotation R are estimated, and the shape adaptation where the shape coefficient vector $\mathbf{b}$ is estimated, are alternatively executed until convergence. For details, we refer to[11] and.[28]

(E) Regularized Shape Deformation: The regularized shape deformation is adapted from[28] whereas for the pelvic and the femoral models instantiated from the last step the correspondences are set up individually but the thin-plate spline (TPS) based shape deformation is solved using the information of both point sets. For details, we refer to.[28]

Implementation Details

RF training was performed only once on a given set of training atlases for 100 pre-defined anatomical landmarks (50 for pelvis and 25 for each proximal femur) and is independent from any target image. The number of sub-volumes used in the training for each atlas is 40,000 while the number of sub-volumes sampled from a given target image in the landmark prediction stage is 20,000. In both training and landmark prediction stages, for each sampled sub-volume, we divided the sub-volume with a grid size $q^3 = 125$. Then for each block, we compute its mean and variance. All the computed means and variances from the 125 blocks are concatenated into a vector which will be used as the feature of the sampled volume.

For the atlases based segmentation, we select $N_s = 2$ atlases, and we use $\lambda = 50$, $\gamma = 0.5$ for graph cut optimization.

Experimental Design

Hip CT data of 30 patients (60 hip joints) were used in our experiments after a local institution review board (IRB) approval. Intra-slice resolutions range from 0.576 mm to 0.744 mm while the inter-slice resolutions have a constant value of 1.6 mm. The number of slices of per CT data ranges from 134 to 206 (on average 159 slices per data). Images are roughly starts from the center of lumbar spine until the lesser trochanter. All CT data are semi-automatically segmented by a trained rater with Amira (www.vsg3d.com/amira).

We designed and conducted a 6-fold cross validation study to evaluate the performance of the present approach. In this study, the 30 CT data was randomly partitioned into 6 equal size subsamples (each subsample has 5 CT data). Of the 6 subsamples, each time a single subsample was used as the test data while the remaining 5 subsamples (the rest 25 CT data) were used as training data. This process was repeated 6 folds, with each one of the 6 subsamples used exactly once as the test data. In each fold, to evaluate the performance of the present approach, we first calculate the segmentation accuracy of the algorithm by comparing the automatic segmentation with the manual segmentation. The widely used measurements of Dice overlap coefficient (DOC)[23] and average surface dis-

TABLE 1. Dice overlap coefficient (DOC) (%) between the automatic segmentation and the manual segmentation when evaluated on 30 hip CT data (60 hips) with a 6-fold cross validation.

Structure	Fold #1	Fold #2	Fold #3	Fold #4	Fold #5	Fold #6	Overall
After multi-atlas based segmentation							
Pelvis	93.6 ± 2.1	94.5 ± 0.7	94.0 ± 1.1	93.1 ± 2.6	94.1 ± 0.9	94.9 ± 1.0	93.9 ± 1.5
Left femur	96.6 ± 0.5	96.8 ± 0.4	96.4 ± 0.9	96.2 ± 0.6	96.6 ± 0.3	96.5 ± 0.5	96.5 ± 0.6
Right femur	96.6 ± 0.2	96.6 ± 0.5	96.3 ± 0.7	95.9 ± 0.3	96.6 ± 0.2	96.6 ± 0.3	96.4 ± 0.5
After aSSM based fitting							
Pelvis	96.0 ± 0.5	96.0 ± 0.5	95.6 ± 0.8	95.5 ± 0.9	95.9 ± 0.7	95.3 ± 0.6	95.7 ± 0.7
Left femur	97.2 ± 0.6	97.4 ± 0.6	97.0 ± 2.1	97.8 ± 0.2	97.3 ± 0.8	97.3 ± 0.6	97.3 ± 1.0
Right femur	97.5 ± 0.3	97.2 ± 0.6	97.2 ± 1.5	97.6 ± 0.5	97.6 ± 0.2	97.0 ± 0.8	97.4 ± 0.7

TABLE 2. Average surface distance (ASD) (mm) between the automatic segmentation and the manual segmentation when evaluated on 30 hip CT data (60 hips) with a 6-fold cross validation.

Structure	Fold #1	Fold #2	Fold #3	Fold #4	Fold #5	Fold #6	Overall
After multi-atlas based segmentation							
Pelvis	0.55 ± 0.11	0.54 ± 0.08	0.54 ± 0.07	0.50 ± 0.15	0.49 ± 0.07	0.49 ± 0.13	0.52 ± 0.10
Left femur	0.46 ± 0.12	0.42 ± 0.08	0.49 ± 0.12	0.48 ± 0.14	0.40 ± 0.07	0.44 ± 0.10	0.45 ± 0.10
Right femur	0.48 ± 0.07	0.47 ± 0.10	0.52 ± 0.07	0.55 ± 0.06	0.44 ± 0.04	0.44 ± 0.08	0.48 ± 0.08
After aSSM based fitting							
Pelvis	0.39 ± 0.09	0.41 ± 0.07	0.40 ± 0.03	0.42 ± 0.13	0.37 ± 0.06	0.41 ± 0.06	0.40 ± 0.07
Left femur	0.40 ± 0.10	0.38 ± 0.09	0.43 ± 0.27	0.29 ± 0.04	0.35 ± 0.11	0.35 ± 0.10	0.36 ± 0.13
Right femur	0.35 ± 0.05	0.41 ± 0.10	0.39 ± 0.20	0.32 ± 0.10	0.30 ± 0.03	0.40 ± 0.12	0.36 ± 0.11

tance (ASD) are selected as the metrics for evaluation. With L_1 being the manual segmentation and L_2 the automatic segmentation, DOC is defined as DOC = $2|L_1 \cap L_2|/(|L_1| + |L_2|)$. To compute the ASD, we first generate a 3D mesh from binary data of each individual segmentation. For each vertex on the surface model derived from the automatic segmentation, we found its shortest distance from the surface model derived from the associated manual segmentation. ASD was then computed as the average of all shortest distances. In order to quantify the contribution of each individual component of the present approach, we computed both metrics in two consecutive stages, i.e., after the multi-atlas based segmentation and after the aSSM based fitting.

In order to investigate the feasibility of using the results from the aSSM based fitting for computer assisted PAO diagnosis and planning, in each fold we further compared differences of three PAO diagnosis parameters, i.e., anteversion,[18] inclination,[18] and AC,[16] when models derived from different segmentation methods were used. All these PAO diagnosis and planning parameters were computed using an in-house developed program.[16]

RESULTS

Quantitative evaluation results for DOC and ASD are presented in Tables 1 and Table 2, respectively. After the multi-atlas based segmentation, a mean DOC of $93.9 \pm 1.5\%$, $96.5 \pm 0.6\%$, and $96.4 \pm 0.5\%$ was found for the pelvis, the left femur and the right femur, respectively, when the results from the multi-atlas based segmentation was compared to the associated manual segmentation. The aSSM based fitting further improved the mean DOC to $95.7 \pm 0.7\%$, $97.3 \pm 1.0\%$, and $97.4 \pm 0.7\%$ for the pelvis, the left femur and the right femur, respectively. When the surface models derived from the multi-atlas based segmentation were compared to those derived from the manual segmentation, a mean ASD of 0.52 ± 0.10 mm, 0.45 ± 0.10 mm and 0.48 ± 0.08 mm was obtained for the pelvis, the left femur, and the right femur, respectively. After the aSSM based fitting, a mean ASD of 0.40 ± 0.07 mm, 0.36 ± 0.13 mm and 0.36 ± 0.11 mm was found for the pelvis, the left femur and the right femur, respectively, which demonstrated that the present approach achieved good segmentation results for all hip bone structures.

In Fig. 5, we visually compare the manual segmentation of a given target image with the result obtained from the multi-atlas based segmentation before aSSM based fitting. Figures 5d and 5g show the color-coded error distributions of the segmented pelvic model and the right femoral model, respectively. It can be seen that overall the segmentation error is small but the resultant surface models are noisy (see Fig. 5f and 5i).

The results from the multi-atlas based segmentation were further refined by the aSSM based fitting. Similar to what we did in Fig. 5, we visually compare the manual segmentation of a given target image with the result obtained from the aSSM based fitting in Fig. 6.
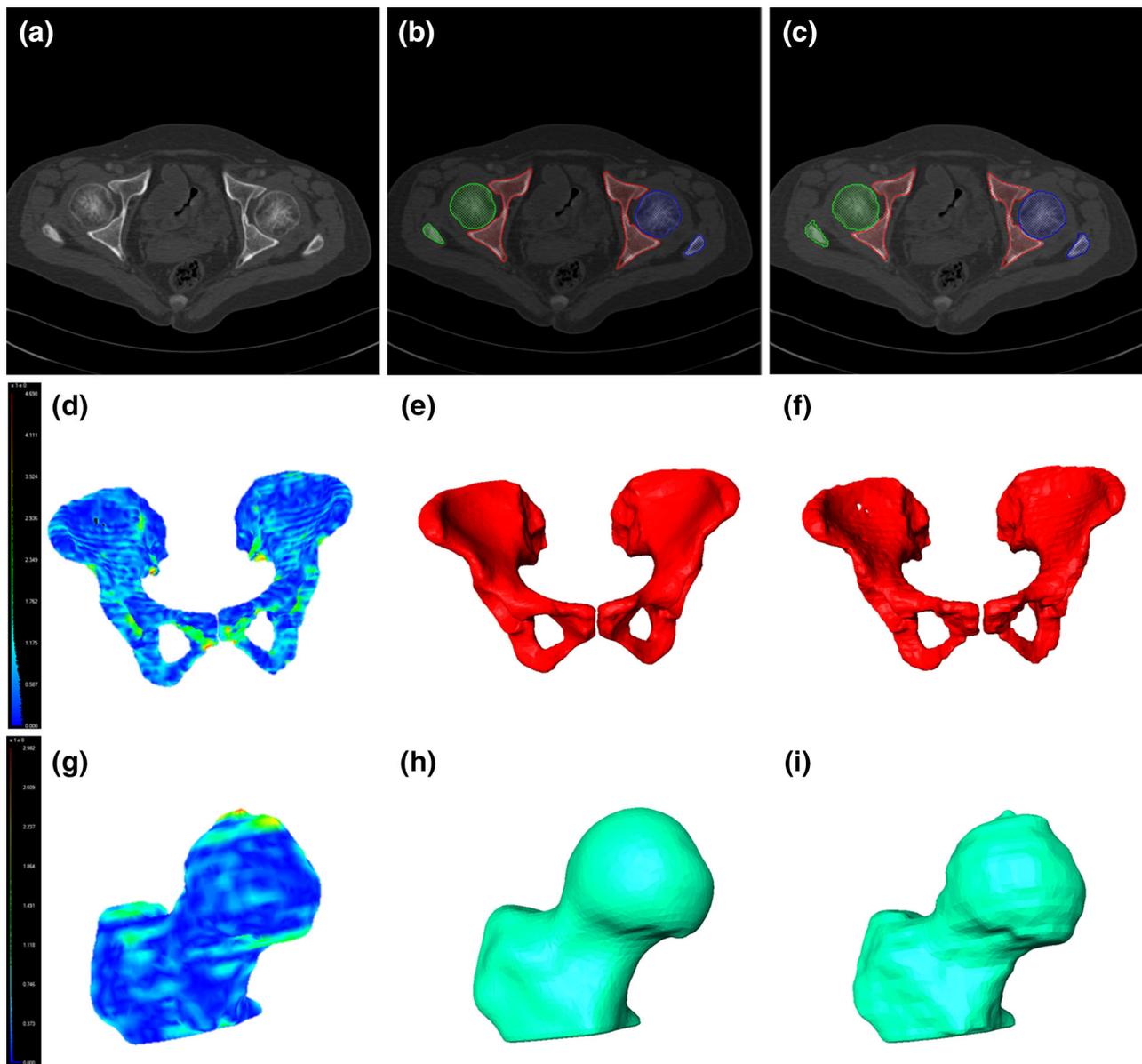
FIGURE 5. Comparison of the results obtained by the multi-atlas based automatic segmentation to manual segmentation of a given target image. (a) An axial slice of the hip joint. (b, e, h) Manual segmentation (b), derived pelvic (e) and femoral (h) models. (c, f, i) the multi-atlas based segmentation (c), derived pelvic (f) and femoral (i) models. Color-coded error distributions of the automatically segmented pelvic surface model (d) and proximal femoral model (g) when compared to associated models derived from manual segmentation.

Figures 6d and 6g show the color-coded error distributions of the segmented pelvic model and the right femoral model, respectively. Comparing the results shown in Fig. 6 with those in Fig. 5, one can clearly see that the resultant surface models in Fig. 6 are visually more similar to the associated ground truth surface models than those in Fig. 5, reflected by the overall decrease of the surface errors.

Figure 7 shows an example of the segmented left hip joint both in 2D axial view and 3D rendering. In order to observe if the joint spaces were preserved and no

penetration of the surfaces was present, we removed parts of the surface and conducted a visual check. From Fig. 7, it can be seen that even the joint space is very narrow, hip joint are accurately reconstructed and there is no penetration between the two surface models.

Quantitative evaluation results for the PAO diagnosis parameters are shown in Table 3. A mean difference of $2.0 \pm 1.5°$ (range from $0.1°$ to $6.4°$), $2.1 \pm 1.6°$ (range from $0.2°$ to $6.7°$) and $3.5 \pm 2.3\%$ (range from 0.1 to 10.1%) were found for anteversion,
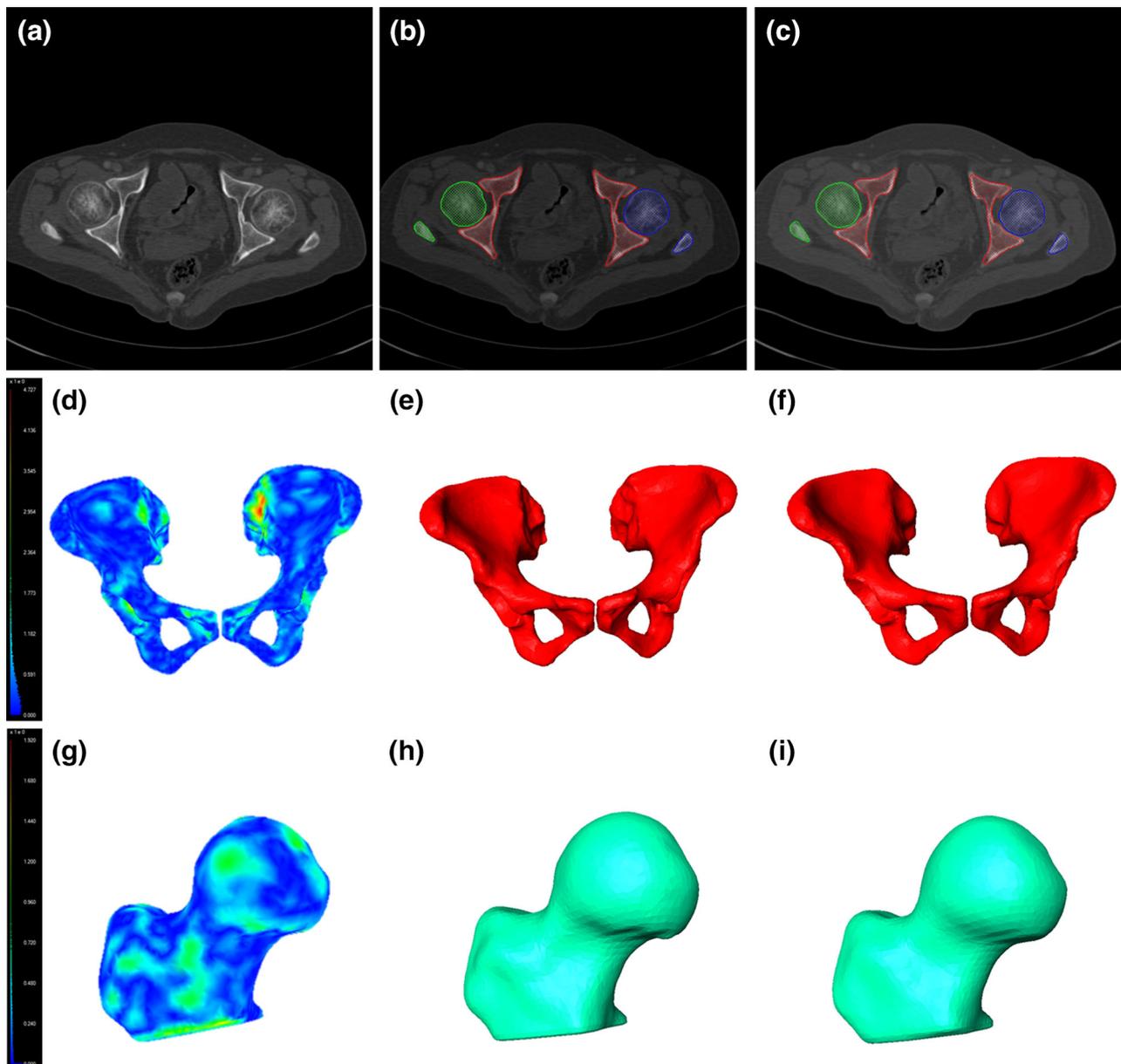
FIGURE 6. Comparison of the results obtained after the aSSM based fitting to manual segmentation of the target image shown in Fig. 5. (a) An axial slice of the hip joint. (b, e, h) Manual segmentation (b), derived pelvic (e) and femoral (h) models. (c, f, i) Segmentation results after aSSM based fitting (c), the fitted pelvic (f) and femoral (i) models. Color-coded error distributions of the automatically extracted pelvic surface model (d) and proximal femoral model (g) when compared to associated models derived from manual segmentation.

inclination, and AC, respectively, when PAO diagnosis parameters derived from the surface models obtained from the aSSM based fitting were compared to those derived from the manual segmentation.

With a single thread-based, unoptimized C++ implementation, on average the run-time of the present approach to segment a hip CT image is 15.8 min, which is about 7.9 min for segmentation of a hip joint on a computer with 3.0 GHz CPU and 12G RAM. As suggested by Cheng et al.,[4] an alternative measurement of the speed of an algorithm is to compute the per slice segmentation time. Our method requires about 6 s per slice.

DISCUSSIONS AND CONCLUSIONS

Accurate extraction of the surface models of a hip joint from CT data is a pre-requisite step for computer assisted diagnosis and planning of PAO surgery. The

goal of the present study is to develop and to validate a fully automatic hip joint segmentation approach. The results of the 6-fold cross validation experiments conducted on 30 CT data (60 hip joints) demonstrated that the present approach was able to accurately segment hip joints. We have also investigated the feasibility of using the automatically segmented models to compute PAO diagnosis parameters such as acetabular orientation (anteversion and inclination) and coverage. The evaluation results demonstrated that the parameters computed from the automatic segmentation were comparable to those computed from the manual segmentation.

The performance of the present approach is compared with the state-of-the-art hip CT segmentation methods.[4,11,14,20,26,27,29] The comparison results are summarized in Table 4, where for comparison purpose, we convert the volumetric overlap error (OE) reported in[11] and[20] to DOC. Details about the conversion can be found in the Appendix.

It is worth to note that due to the fact that different datasets are used in evaluation of different methods, direct comparison of different methods is difficult. Thus, the comparison results in Table 4 should be interpreted cautiously, even though here we only compared the results obtained from the dataset which are similar to ours (patient data with normal or mildly diseased hip joints). Nevertheless,
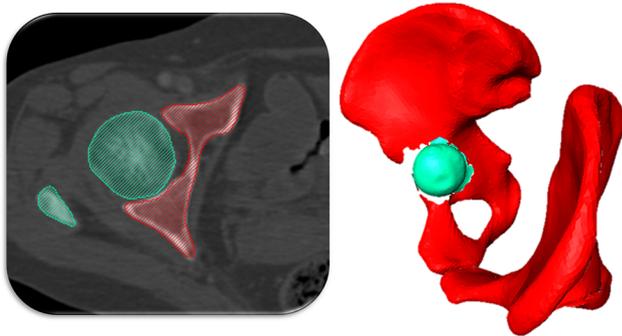

FIGURE 7. The segmented models of pelvis (red) and left proximal femur (cyan). Left: a 2D axial slice. Right: 3D surface rendering. Part of the pelvic surface model is removed for visualizing the reconstructed hip joint surface models. No penetration between models of the neighboring structures is found.

as shown in Table 4, our approach is comparable to most of the existing work including the intensity based methods[4,29] and SSM based methods.[14,20,26,27] One possible explanation why the present approach achieves better results than the existing intensity based methods[4,29] is that due to the integration of prior knowledge from atlases and aSSMs, regions which are difficult to be segmented using only intensity information could be accurately segmented by our approach. In comparison to the SSM based method,[14,20,26,27] our approach also achieved superior or equivalent results. One possible explanation is that SSM based methods usually depend on a good shape model. Thus the SSM fitting quality is variable. In contrast, the present approach integrates prior knowledge from atlases selected after a landmark based registration and aSSMs, leading to not only a more robust and accurate segmentation but also a preservation of the hip joint structure. This has been demonstrated in Figs. 4 and 7. Our finding goes along with the results in[11] where an equivalent accuracy to our approach (i.e., the method in[11] achieved a slightly better ASD but a slightly worse average DOC than our approach) was reported. The differences between our approach and the method reported in [11] are: (a) we use a fast RF regression based landmark detection to initialize the segmentation process while the method introduced in[11] is based on a generalized Hough transform; and (b) we combine an atlas based segmentation method with an aSSM-based fitting approach for a robust and accurate segmentation while their method is based on the image driven aSSM adaptation, solved with a graph based optimization.[11]

From Table 4, one can see that from a per CT stack or per hip computational time point of view our approach is slower than the methods reported in,[11,20] and[4] but faster than other methods.[26,27,29] Again, such a comparison should be interpreted cautiously due to the fact that different datasets with different resolutions were used to evaluate different methods. An alternative measurement of the speed of an algorithm, as suggested by Cheng *et al.*,[4] is to compute the per slice segmentation time. If such a metric is used, the speed of the present approach is comparable to most of the state-of-the-art methods.[4,26,27,29]

TABLE 3. PAO diagnosis and planning parameter differences when surface models derived from different segmentation methods were used.

Parameters	Manual segmentation	Automatic segmentation	Mean absolute differences
Anteversion (°)	13.7 ± 4.8	13.5 ± 4.7	2.0 ± 1.5 (range from 0.1–6.4)
Inclination (°)	49.2 ± 4.2	49.1 ± 3.0	2.1 ± 1.6 (range from 0.2–6.7)
Coverage (%)	82.2 ± 7.3	79.5 ± 5.7	3.5 ± 2.3 (range from 0.1–10.1)

TABLE 4. Comparison of the results achieved by the present method with those reported in the literature.

Methods	Segmented structure	Preserving hip joint	ASD (mm)	Average DOC (%)	Average time	Seconds/slice
Lamecker et al.[14]	Complete pelvis	No	1.80	—	—	—
Seim et al.[20]	Complete pelvis	No	0.40 (RHB), 0.60 (LHB)	93.67 (RHB), 91.54 (LHB)	4.2 min per CT Stack	—
Kainmueller et al.[11]	Complete pelvis and one-side femur	Yes	0.30 (RHB), 0.60 (LHB)	94.90 (RHB), 92.01 (LHB)	6.0 min per CT stack	—
Cheng et al.[4]	Partial pelvis and both femurs	Yes	0.86	94.44	5.9 min per hip	8
Yokota et al.[27]	Partial pelvis and both femurs	No	1.10	92.69	10.1 min per hip	13.6
Zoroofi et al.[29]	Partial pelvis and both femurs	No	0.91	93.89	7.0 min per hip	9.5
Yokota et al.[26]	Complete pelvis and both femurs	No	0.98	—	15.0 min per hip	—
Our method	Complete pelvis and both femurs	Yes	0.37	96.8	7.9 min per hip	6

The cited results of[27] and[29] were extracted from.[4] Since the authors in[4] evaluated all three algorithms on four datasets, here we only took the best results achieved by each algorithm. For[11] and,[20] we cited the results for segmenting both the right hip bone (RHB) and the left hip bone (LHB) from 50 CT stacks. With a voxel size of about $0.9 \times 0.9 \times 5.0$ mm^3, all these 50 CT stacks approximately have the same resolution.[20] Please note, each CT stack included an implant on either the left or right acetabulum. At the prosthesis side, the implant itself was not included in the segmentation. Aiming for an independent evaluation, all datasets with the implant on the right were mirrored before segmentation such that all had the native hip joint on the right and the implant on the left side.[20] Furthermore, for comparison purpose, we converted the Volumetric Overlap Error (OE) reported in[11] and[20] to Dice Overlap Coefficient (DOC). See text for details.

It is worth to mention the limitations of the present study. First, the segmentation performance of our approach is highly dependent on the atlas based segmentation, which uses a discrete optimization based non-rigid registration[9] to register selected atlases to the target image. Although our approach is robust to align atlases to target image space globally, there might be mis-registration in small local areas at the corner or along the boundary of the bone. This can be observed in Fig. 6 where the small area around the left sacroiliac joint has relatively large errors. One possible solution is to use local registration to improve the registration performance, which has been proved to be useful for soft tissue segmentation.[5] Second, the present approach was evaluated on limited patient data (60 hips) with normal or mildly diseased hip joints. Nonetheless, the experimental results from the 6-fold cross validation study demonstrated the accuracy and the robustness of the present approach.

In summary, we presented a fully automatic method for segmenting CT images of a hip joint for computer assisted diagnosis and planning of PAO surgery. The strength of the present approach lies in the combination of a fast RF regression based landmark detection, atlas-base segmentation with an aSSM based model instantiation. The present approach can be extended to segment CT data of other anatomical structures.

APPENDIX

How to convert Volumetric Overlap Error (OE) to Dice Overlap Coefficient (DOC)?

Let L_1, $L_2 \subset \mathbb{R}^3$ denote the ground truth segmentation and the automatic segmentation, OE is calculated with the definition described in[30]:

$$\text{OE} = 1 - \frac{(|L_1 \cap L_2|)}{(|L_1 \cup L_2|)} \tag{9}$$

And DOC is defined as

$$\text{DOC} = 2 \times \frac{|L_1 \cap L_2|}{(|L_1| + |L_2|)} \tag{10}$$

Since we have

$$(|L_1 \cup L_2|) = (|L_1| + |L_2|) - |L_1 \cap L_2| \tag{11}$$

After some mathematic manipulations, one can derive following relationship between OE and DOC:

$$\text{DOC} = 2 \times \frac{(1.0 - \text{OE})}{(2.0 - \text{OE})} \tag{12}$$

For example, in Ref.,[11] Kainmueller et al. reported an average OE of 9.7% for segmenting right hip bone

(RHB) (Table IV of Ref. 11). The corresponding DOC is then 94.90%.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

The authors have no conflict of interest related to this work.

REFERENCES

[1]Aljabar, P., R. Heckemann, A. Hammers, J. Hajnal, and D. Rueckert. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuro-Image* 46(3):726–738, 2009.

[2]Boykov, Y., O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* 23(11):1222–1239, 2001.

[3]Breiman, L. Random forests. *Mach. Learn.* 45(1):5–32, 2001.

[4]Cheng, Y., S. Zhou, Y. Wang, C. Guo, J. Bai, and S. Tamura. Automatic segmentation technique for acetabulum and femoral head in CT images. *Pattern Recogn.* 46(11):2969–2984, 2013.

[5]Chu, C., M. Oda, T. Kitasaka, K. Misawa, M. Fujiwara, Y. Hayashi, Y. Nimura, D. Rueckert, and K. Mori. Multi-organ segmentation based on spatially-divided probabilistic atlas from 3D abdominal CT images. MICCAI 2013, Part II, pp. 165–172, 2013.

[6]Criminisi, A., J. Shotton, D. Robertson, and E. Konukoglu. Regression forests for efficient anatomy detection and localization in CT studies. *MCV* 2010:106–117, 2010.

[7]Ehrhardte, J., H. Handels, W. Plotz, and S. J. Poppl. Atlas-based recognition of anatomical structures and landmarks and the automatic computation of orthopedic parameters. *Methods Inf. Med.* 43(4):391–397, 2004.

[8]Ganz, R., K. Klaue, T. Vinh, and J. Mast. A new periacetabular osteotomy for the treatment of hip dysplasia: technique and preliminary results. *Clin. Orthop.* 232:26–36, 1988.

[9]Glocker, B., N. Komodakis, G. Tziritas, N. Navab, and N. Paragios. Dense image registration through MRFs and efficient linear programming. *Med. Image Anal.* 12(6):731–741, 2008.

[10]Haas, B., T. Coradi, M. Scholz, P. Kunz, M. Huber, U. Oppitz, L. André, V. Lengkeek, D. Huyskens, A. van Esch, and R. Reddick. Automatic segmentation of thoracic and pelvic CT images for radiotherapy planning using implicit anatomic knowledge and organ-specific segmentation strategies. *Phys. Med. Biol.* 53(6):1751–1771, 2008.

[11]Kainmueller, D., H. Lamecker, S. Zachow, and H.-C. Hege. An articulated statistical shape model for accurate hip joint segmentation. IEEE EMBC, pp. 6345–6351, 2009.

[12]Kainmueller, D., H. Lamecker, S. Zachow, and H. C. Hege. Coupling deformable models for multi-object segmentation. *ISBMS* 2008:69–78, 2008.

[13]Kang, Y. K. Engelke, and W. and A. Kalender. A new accurate and precise 3D segmentation method for skeletal structures in volumetric CT data. *IEEE Trans. Med. Imaging* 22(5):586–598, 2003.

[14]Lamecker, H., M. Seebaß, H. C. Hege, and P. Deuflhard. A 3D statistical shape model of the pelvic bone for segmentation. *SPIE* 5370:1341–1351, 2004.

[15]Lindner, C. S. Thiagarajah, J. M. Wilkinson, arcOGEN Consortium, G. Wallis, and T. F. Cootes. Fully automatic segmentation of the proximal femur using random forest regression voting. *IEEE Trans. Med. Imaging* 32(8):1462–1472, 2013.

[16]Liu, L., T. Ecker, S. Schumann, K. Siebenrock, L.-P. Nolte, and G. Zheng. A novel planning and navigation system for peri-acetabular osteotomy (PAO). *Int. J. Comput. Assist. Radiol. Surg.* 9(Suppl 1):S168–S169, 2014.

[17]McKinley, T. O. The Bernese Periacetabular Osteotomy: review of Reported Outcomes and the Early Experience at the University of Iowa. *Iowa Orthop. J.* 23:23–28, 2003.

[18]Murray, D. W. The definition and measurement of acetabular orientation. *J. Bone Joint Surg. [Br.]* 75-B:228–232, 1993.

[19]Pettersson, J., H. Knutsson, and M. Borga. Automatic Hip Bone Segmentation Using Non-Rigid Registration. *ICPR* 2006:946–949, 2006.

[20]Seim, H., D. Kainmueller, M. Heller, H. Lamecker, S. Zachow, and H. C. Hege. Automatic segmentation of the pelvic bones from CT data based on a statistical shape model. *VCBM* 2008:93–100, 2008.

[21]Vercauteren, T., X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: efficient non-parametric image registration. *NeuroImage* 45(1):61–72, 2009.

[22]Viola, P., and M. Jones. Rapid object detection using a boosted cascade of simple features. CVPR 2001, vol. I, pp. 511–518, 2001.

[23]Wolz, R., C. Chu, K. Misawa, M. Fujiwara, K. Mori, and D. Rueckert. Automated abdominal multi-organ segmentation with subject-specific atlas generation. *IEEE Trans. Med. Imaging* 32(9):1723–1730, 2013.

[24]Xia, Y., J. Fripp, S. S Chandra, R. Schwarz, C. Engstrom and S. Crozier. Automated bone segmentation from large field of view 3D MR images of the hip joint. *Phys. Med. Biol.* 58:7375–7390.

[25]Yang, C., R. Duraiswami, and L. Davis. Efficient Kernel Machines Using the Improved Fast Gauss Transform. *NIPS* 2005:1561–1568, 2005.

[26]Yokota, F., T. Okada, M. Takao, S. Sugano, Y. Tada, N. Tomiyama, and Y. Sato. Automated CT segmentation of diseased hip using hierarchical and conditional statistical shape models. *MICCAI 2013, Part II*, pp. 190–197, 2013.

[27]Yokota, F., T. Okada, M. Takao, S. Sugano, Y. Tada, and Y. Sato. Automated segmentation of the femur and pelvis from 3D CT data of diseased hip using hierarchical statistical shape model of joint structure. *MICCAI 2009, Part II*, pp. 811–818, 2009.

[28]Zheng, G., X. Dong, K. Rajamani, X. Zhang, M. Styner, R. Thoranaghatte, and L.-P. Nolte. MA Gonzalez Ballester. Accurate and robust reconstruction of a surface model of the proximal femur from sparse-point data and a dense-point distribution model for surgical navigation. *IEEE Trans. Biomed. Eng.* 54:2109–2122, 2007.

[29]Zoroofi, R. A., Y. Sato, T. Sasama, T. Nishii, N. Sugano, K. Yonenobu, H. Yoshikawa, T. Ochi, and S. Tamura. Automated segmentation of acetabulum and femoral head from 3-D CT images. *IEEE Trans. Inf. Technol. Biomed.* 7(4):329–343, 2003.

[30]van Ginneken, B., T. Heimann, and M. Styner. 3D Segmentation in the Clinic: A Grand Challenge. In: T. Heimann, M. Styner, B. van Ginneken (Eds.): 3D Segmentation in the Clinic: A Grand Challenge, pp. 7-15, 2007.