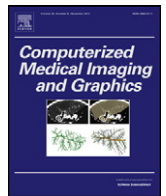




Contents lists available at SciVerse ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag



Modeling of the bony pelvis from MRI using a multi-atlas AE-SDM for registration and tracking in image-guided robotic prostatectomy

Qinquan Gao^a, Ping-Lin Chang^a, Daniel Rueckert^a, S. Mohammed Ali^b, Daniel Cohen^c, Philip Pratt^d, Erik Mayer^c, Guang-Zhong Yang^d, Ara Darzi^c, Philip “Eddie” Edwards^{c,*}

^a Department of Computing, Imperial College London, UK

^b Department of Colorectal Surgery, The Royal Marsden Hospital NHS Trust, London, UK

^c Department of Surgery and Cancer, Imperial College London, UK

^d Hamlyn Centre for Robotic Surgery, Imperial College London, UK

ARTICLE INFO

Keywords:

Statistic deformation model
Multi-atlas segmentation
Hierarchical registration
Multi-modality registration
Augmented reality
Dense tracking and mapping

ABSTRACT

A fundamental challenge in the development of image-guided surgical systems is alignment of the preoperative model to the operative view of the patient. This is achieved by finding corresponding structures in the preoperative scans and on the live surgical scene. In robot-assisted laparoscopic prostatectomy (RALP), the most readily visible structure is the bone of the pelvic rim. Magnetic resonance imaging (MRI) is the modality of choice for prostate cancer detection and staging, but extraction of bone from MRI is difficult and very time consuming to achieve manually. We present a robust and fully automated multi-atlas pipeline for bony pelvis segmentation from MRI, using a MRI appearance embedding statistical deformation model (AE-SDM). The statistical deformation model is built using the node positions of deformations obtained from hierarchical registrations of full pelvis CT images. For datasets with corresponding CT and MRI images, we can transform the MRI into CT SDM space. MRI appearance can then be used to improve the combined MRI/CT atlas to MRI registration using SDM constraints. We can use this model to segment the bony pelvis in a new MRI image where there is no CT available. A multi-atlas segmentation algorithm is introduced which incorporates MRI AE-SDMs guidance. We evaluated the method on 19 subjects with corresponding MRI and manually segmented CT datasets by performing a leave-one-out study. Several metrics are used to quantify the overlap between the automatic and manual segmentations. Compared to the manual gold standard segmentations, our robust segmentation method produced an average surface distance 1.24 ± 0.27 mm, which outperforms state-of-the-art algorithms for MRI bony pelvis segmentation. We also show that the resulting surface can be tracked in the endoscopic view in near real time using dense visual tracking methods. Results are presented on a simulation and a real clinical RALP case. Tracking is accurate to 0.13 mm over 700 frames compared to a manually segmented surface. Our method provides a realistic and robust framework for intraoperative alignment of a bony pelvis model from diagnostic quality MRI images to the endoscopic view.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Prostate cancer is the second most frequently diagnosed cancer in males globally, and the third leading cause of cancer death in males in the developed world [1]. Radical prostatectomy is an effective surgical treatment for organ-confined disease. However there is a clinical need to improve functional outcomes, such as continence and potency, together with oncological control [2]. Robot-assisted laparoscopic prostatectomy enables the operating surgeon to view the operating field in stereo, offering the possibility of incorporating intraoperative 3D image guidance with

augmented reality, which has the potential to improve patient outcomes by improving surgical accuracy and decreasing the learning curve [3].

To achieve such image guidance, a 3D model from preoperative imaging must be aligned with the view of the patient through the stereo endoscope. Construction of the 3D model of the anatomy is the first requirement. Manual segmentation by a specialist radiologist is considered to be the only reliable gold standard for delineating anatomical structures. This process is particularly time consuming and is a barrier to routine clinical implementation of image guidance. MRI is becoming the gold standard imaging modality for detection, localisation and staging of prostate cancer by providing detailed functional and anatomical tissue information [4]. Segmentation of the large bony pelvis from MRI is particularly difficult even for experienced radiologists, as the

* Corresponding author. Tel.: +44 7971 681122.

E-mail address: eddie.edwards@imperial.ac.uk (P. Edwards).

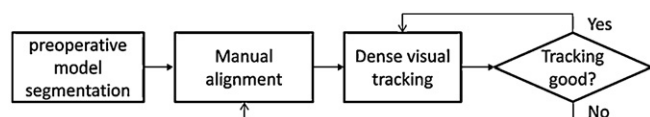


Fig. 1. The framework of proposed system for automated identification of the bony pelvis from MRI and tracking of this registered model in the surgical scene.

cortical bone is dark and can be hard to distinguish from surrounding tissue. Diagnostic MRI scans are typically not high resolution, having a slice separation of several millimeters. CT would provide good bone delineation, but it is hard to justify the additional costs and radiation dose involved since CT does not provide clear soft tissue delineation. Therefore, automatic and accurate segmentation from diagnostic quality MRI scans is desirable and presents a research challenge.

During RALP, pneumoperitoneum and tissue dissection causes significant soft tissue deformation. Therefore, the large and rigid bony pelvis presents the most sensible target for intraoperative rigid registration and tracking. The pelvic rim is visible in the intraoperative view through the stereo laparoscope. The segmented bony pelvis can be manually aligned to the view through the da Vinci stereo-endoscope and subsequent tracking of the motion of the endoscopic camera can be achieved using dense visual tracking methods [5]. The result enables near real-time alignment of the preoperative model and can provide augmented reality guidance to the surgeon. Fig. 1 provides a flow diagram of our system.

2. Background to the proposed method

Several algorithms have been proposed to segment bone from MRI Images. Lorigo et al. [6] incorporated texture information with a geodesic active contours framework to automatically detect closed boundaries in 2D MRI of the knee. Rifa et al. [7] used deformation models, which take into account partial volume effects, to segment the skull bone in MRI volumes. Schmid et al. [8] combined the physically-based deformable models with shape priors to automatically segment femur and hip bones in low resolution clinical MRI scans. Schmid et al. [9] further presented a robust multi-resolution SSM algorithm with an adapted initialization to segment MRI femur and hip bone images acquired in small field of views.

In our specific task of segmenting the bony pelvis from MRI for image-guided prostate surgery, the accuracy of the bony pelvis segmentation will significantly affect the accuracy of subsequent registration and tracking during surgery. Dense tracking algorithms aim to calculate the camera pose in successive frames of a video by back projecting the texture from one frame onto the aligned model and then projecting this into subsequent frames. Pixel-wise similarity is optimized to find the camera pose [5]. These methods will fail if they are directly applied to inaccurately segmented bone structures. Segmentation of bone from MRI images is difficult since cortical bone is dark in MRI and can be difficult to distinguish from surrounding structures. Most of the existing algorithms for bony pelvis segmentation are based on CT images [10–12]. In order to tackle this problem, Thompson et al. [13] used a statistical deformation model constructed from CT images coupled to MRI appearance data for MRI bony pelvis segmentation. This is the most closely related work to that presented here. Results are limited to a single atlas-based segmentation [14], but the method shows some promise.

A promising approach to improve segmentation accuracy is to combine basic atlas-based segmentation with classifier fusion, a technique from machine learning [15]. In this approach several atlases from different subjects are registered to target data. The label that the majority of all warped labels predict for each voxel

is used for the final segmentation of the target image. Such a multi-atlas segmentation provides the best accuracy compared to a number of algorithms for the segmentation of subcortical structures in a recent study [16]. To the knowledge of the authors the multi-atlas method has not been applied to bony pelvis segmentation.

Direct multi-atlas-based segmentation uses free-form deformations that can potentially result in local minima that correspond to unrealistic deformations. Statistical atlas-based methods are popular in many medical image segmentation tasks, since they are capable of handling shapes with large but regular variation and provide a comprehensive description of the underlying shape variation when the training data is sufficient [17]. In this paper, we propose a multi-atlas segmentation of MRI scan of the bony pelvis by combining shape information of the pelvis derived from CT data with appearance information from an MRI volume (AE-SDMs).

To construct a shape model of pelvic anatomical variation for each atlas, we initially used 30 CT scans of the full pelvis. Using the CT scan of the atlas as a template, all the CT scans in the training set were accurately registered using manually identified landmarks and segmented cortical bone surfaces. A symmetric cost function involving surface and landmark distance provides the most robust registration. A subsequent intensity-based registration provides a B-spline non-rigid transformation using the method of Rueckert et al. [18]. All combinations of CT training scans were registered in this way and a statistical deformation model (SDM) [19,20] was constructed with each CT scan as a template for each atlas.

For 19 of the CT scans in the training set, we have corresponding MRI scans (see Fig. 2). Though these have lower resolution in the z-plane, we can accurately register these to the corresponding CT scan using a 9 degree-of-freedom transformation incorporating rigid motion and anisotropic scaling along the axes. This provides an MRI intensity appearance embedded within the SDM space to create an AE-SDM.

To fit the model to an unseen MRI scan, we use MRI intensities from the AE-SDM and normalized mutual information as a similarity measure. This is done for each of the atlases in turn. A coarse SDM registration is followed by a free-form deformation. The atlases are first classified according to subpubic angle and normalized mutual information between the atlas image and the template image. The most similar N atlases (usually 10–12) between the deformed atlases and the target are selected [21]. For these improved transformations, the bony pelvis labels of the atlas are propagated into the space of the target image. The label that the majority of all warped labels predict for each voxel is used for the final segmentation of the target image.

Finally, we align this automatically segmented model to the view through the da Vinci stereo-endoscope. Initial registration is manual, following a protocol whereby an initial stereo landmark is fixed and rotation is performed around this point. Such a protocol has been suggested for partial nephrectomy [22]. Subsequent motion of the endoscopic camera is tracked using dense tracking methods [23].

We begin by describing each of the stages of this process and present results in terms of both segmentation and tracking accuracy with a final demonstration on a real clinical dataset captured during RALP.

3. Multi-SDM construction with MRI appearance embedding (AE-SDMs)

A statistical deformable atlas consists of a reference image coupled with its segmented labels and a statistical deformation model. This model, which consists of the mean deformation and the principle modes of variation from the mean deformation field,

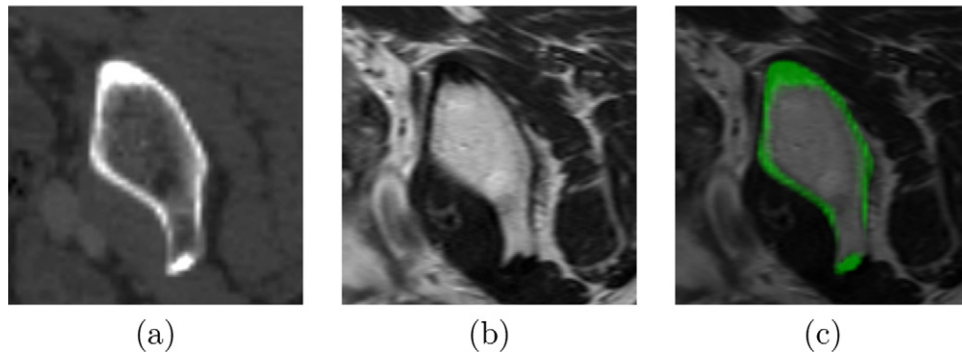


Fig. 2. An example CT (a) and MRI scan (b) with accurate registration (c).

describes how the atlas should be deformed to fit to a new given target image. CT images represent high resolution bone information. Hence, good mapping between two different individuals can be obtained using non-rigid registration. However, direct alignment of a CT atlas to a target MRI image performed poorly in our experiments, since the intensity of cortical bone is dark in MRI and can be hard to distinguish from surrounding tissue. In order to tackle this problem we use a number of training examples with corresponding CT and MRI images of the same patient. The MRI image can be accurately registered to CT space to construct an MRI appearance embedding statistical deformation model (AE-SDM), which can guide non-rigid registration to establish good alignments with a new MRI image.

For the purpose of multi-atlas segmentation with AE-SDMs guidance, we carried out a set of registrations using each of the atlas CT images as the template, yielding the deformations with all other CT images. A separate SDM is then built for each atlas. A hierarchical registration scheme is proposed to produce more accurate transformations, consisting of landmark and surface guided registration, intensity based affine registration and non-rigid registration using B-spline based free-form transformation [24,18]. The non-rigid deformations are used to build the SDM using principle component analysis. The corresponding MRI scan is warped into the SDM space to construct the AE-SDM. Fig. 3 presents the whole process of the AE-SDMs construction.

3.1. Non-rigid alignment of the CT bone models using hierarchical registration

To build a statistical deformation model, it is vital that the training registrations are as accurate as possible. With this aim, two phases of registrations from coarse to fine are implemented. In the first phase, an initial 47 manually defined landmarks covering the whole area of a pelvis were introduced and landmark-based affine registration was carried out (see Fig. 4). This was followed by intensity-based 9 degree-of-freedom registration (rigid plus anisotropic scaling) on the bony pelvis masks with 3 mm dilation. This allows the outer bone edge to be included, but avoids soft tissue affecting the registration. Subsequently, a non-rigid surface-based registration with landmark-constraint is proposed to further improve the alignment. The symmetric surface distance and landmark distance constraint are used as the cost function, which can be denoted as:

$$S = \frac{(D(S_{R,i}, T(S_{A,i})) + D(S_{A,i}, T^{-1}(S_{R,i})))}{2} + \lambda D(L_{R,j}, T(L_{A,j})) \quad (1)$$

where D is root mean square distance, $S_{R,i}$ is surface points of reference atlas, $S_{A,i}$ is surface points of target images, $L_{R,j}$ and $L_{A,j}$ are landmarks, and λ is a trade-off between the two terms. The value 4 was found to be optimal during our experiments (see Fig. 4).

In the second phase, a three-level free-form non-rigid registration is introduced to further improve the alignment. The masked bony pelvis with 20 mm dilation is used in both the reference atlas and the target image to enable a small amount of surrounding soft tissue to influence the registration. In this process, we chose normalized mutual information (NMI) as the similarity measure; B-spline transformation combined with affine transformation as the transformation and gradient descent as the optimization method. Registration is carried out based in a hierarchical manner from low resolution to high resolution with different control point spacing (30 mm, 15 mm, 7.5 mm). The output local transformation from each level represents a 3D deformation field of control points showing how the deformations required to transform the reference atlas to the target, denoted as:

$$T_{x,y}^{FFD} = \begin{pmatrix} 0 & T_{1,2}^{FFD} & T_{1,3}^{FFD} & \dots & T_{1,n}^{FFD} & \dots & T_{1,m}^{FFD} \\ T_{2,1}^{FFD} & 0 & T_{2,3}^{FFD} & \dots & T_{2,n}^{FFD} & \dots & T_{2,m}^{FFD} \\ T_{3,1}^{FFD} & T_{3,2}^{FFD} & 0 & \dots & T_{3,n}^{FFD} & \dots & T_{3,m}^{FFD} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \dots & \vdots \\ T_{n,1}^{FFD} & T_{n,2}^{FFD} & T_{n,3}^{FFD} & \dots & 0 & \dots & T_{n,m}^{FFD} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ T_{m,1}^{FFD} & T_{m,2}^{FFD} & T_{m,3}^{FFD} & \dots & T_{m,n}^{FFD} & \dots & 0 \end{pmatrix} \quad (2)$$

where $T_{x,y}^{FFD}$ is a non-rigid transformation from atlas x to atlas y using free from deformation (FFD), $T_{x,y}^{FFD} = T_{y,x}^{FFD^{-1}}$ and when $x > n$, $T_{x,y}^{FFD} = 0$, n represents the whole number of atlases in the pool, and m is the whole number of CT scans. All training CT datasets were registered to all other datasets.

3.2. Principal component analysis of CT deformation model

For easier implementation, principle component analysis was performed on the deformation field $T_{x,y}^{FFD}$ to build up the statistical deformation model for each reference atlas. Only 29 transformations were used to build the SDM, and the remaining dataset is used to validate this model in a leave-one-out manner. These models capture the likely range of pelvic shape and can be used to improve the registration from the atlas image to a new target by reducing search space of the non-rigid registration. Though a larger training set would be desirable, the outcome of 29 modes can approximately represent the whole variance space. The cumulative percentage of population variance is shown in Fig. 5. The first 15 modes account for 91.07% of the variance.

Varying the 3 leading modes of SDM parameters between standard deviation from mean value ($V = \pm 3$), the variation of free form and the effects on the underlying images are shown in Fig. 6.

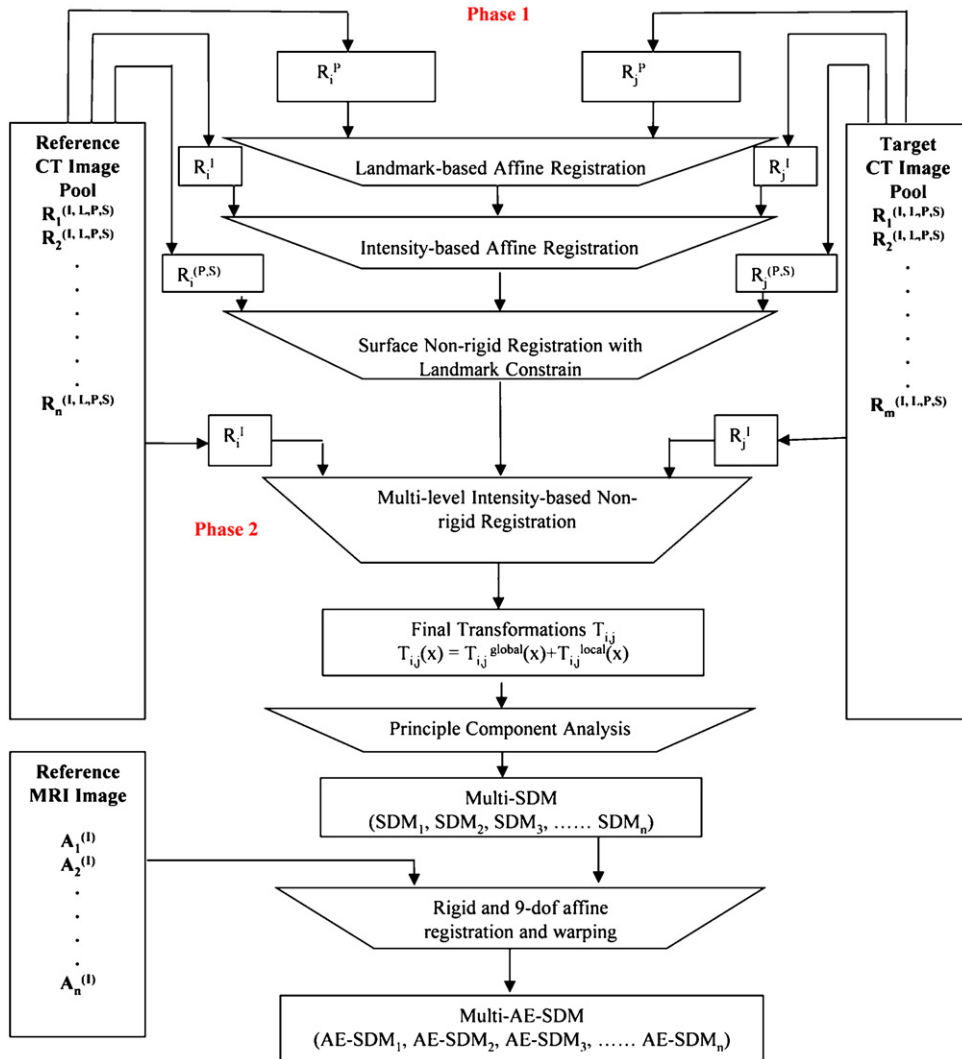


Fig. 3. AE-SDMs building process in different coordinate: in each CT image $R_n^{(I,L,P,S)}$, I represents the intensity image of CT scan, L is a segmentation of bony pelvis of the CT scan, P is a landmark set and S is a bony pelvis surface correspondingly, $n \leq m$. In each MRI image $A_n^{(I)}$, I represents the intensity image of MRI scan.

3.3. MRI appearance embedding SDMs (AE-SDMs)

In order to incorporate MRI intensity information for the registration from atlas to a new MRI image, we can register the corresponding MRI images into each atlas CT image space (see

Fig. 2). First, each MRI image is transformed to its corresponding masked CT image using a rigid transformation, followed by a 9-dof affine registration since there maybe slight scaling differences between scanners. In some cases a rough initial manual alignment is required. Subsequent rigid and affine registration are used to

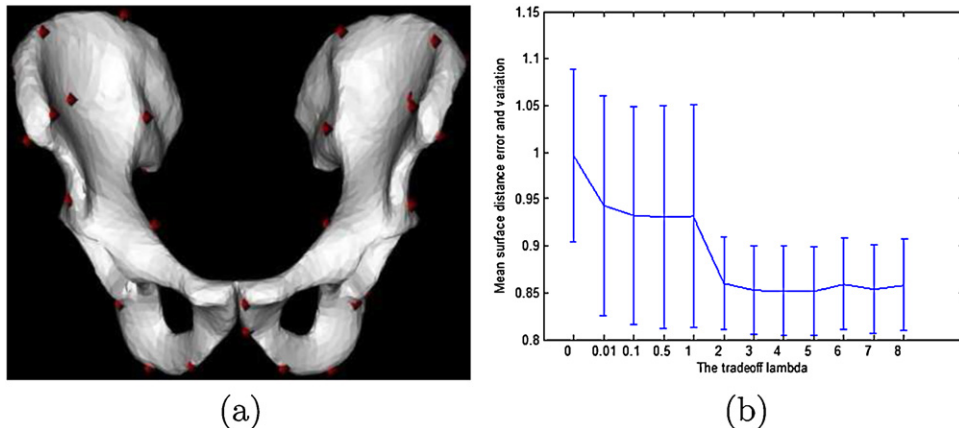


Fig. 4. Chosen surface landmarks (a) and optimization of surface vs landmark parameter, (b).

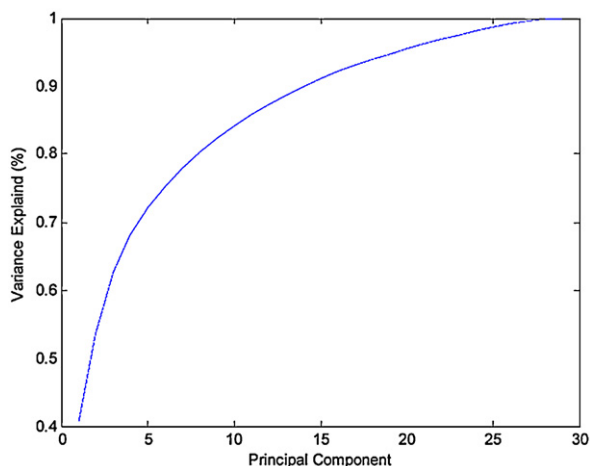


Fig. 5. Cumulative percentage of population variance.

further improve the transformation $T_j^{MRI \rightarrow CT}$. By applying the $T_{ij}(x)$ (showed in Fig. 3) from the target to reference atlas transformation and the transformation from MRI to CT $T_j^{MRI \rightarrow CT}$, MRI appearance could be generated in each CT atlas space, denotes as:

$$X_{MRI_j}^{Atlas_i} = T_{i,j}(x) T_j^{MRI \rightarrow CT} X_{MRI_j} \quad (3)$$

Then, an average MRI appearance could be generate for each atlas, represented as:

$$A_i^{MRI} = \frac{\sum_{j=0}^n X_{MRI_j}^{Atlas_i}}{n} \quad (4)$$

In practice, there was little or no performance improvement using an average MRI compared to the rigidly registered single MRI scan for the given atlas.

The result of the above processes is a series of AE-SDM models with an MRI intensity model and the corresponding shape variation model to describe how to warp this template to a new MRI scan.

4. Multi-atlas segmentation of bony pelvis from MRI using AE-SDMs

The next step was to utilise the multiple AE-SDMs to fit to an unseen MRI, providing an accurate segmentation of the bone. In multi-atlas-based segmentation, the intensity images of selected atlas are registered non-rigidly to a target image and the resulting transformation is used to propagate the anatomical structure labels of the atlas into the space of the target image. The label that the majority of all warped labels predict for each voxel is used for the final segmentation of the target image. In our case there are only two labels, bone and background. The pipeline of multi-atlas segmentation of bony pelvis is divided into the following parts: atlas selection, AE-SDM-guided hybrid registration, and atlas propagation and fusion, as shown in Fig. 7.

4.1. Atlas selection using shape and intensity similarity

The best multi-atlas segmentation accuracy is obtained with a subset of the atlases instead of choosing all the atlases. For each new query image to be segmented, the atlases in the database that are the most similar to the query image should be used [25,26]. Several methods have been proposed and compared to improve this selection by Lotjonen [15]. Since the pelvic shapes between male and female are quite different, it is better to classify the atlases into different shape categories according to the gender indicators. The sub-pubic angle (denoted as θ_p) is important in forensic anthropology in determining the sex of an unknown person from skeletal remains. Generally, the sub-pubic angle is significantly wider in women than men [27]. In most cases, a sub-pubic angle of 50–82 deg indicates a male with smaller pelvic shape; an angle of >90 deg indicates a female with larger pelvic shape [28]. Between these two indicators, a uncertain case is defined with middle size of pelvic shape. Therefore, in the first step, an automatic sub-pubic angle measurement of both CT and MRI scans is introduced as the gender classifier for coarse atlas selection. The sub-pubic angle can be estimated by applying the non-rigid transformation to the three relevant anatomical points from the template. Then, a classifier (Eq. (5)) is used to divide the atlas into three categories:

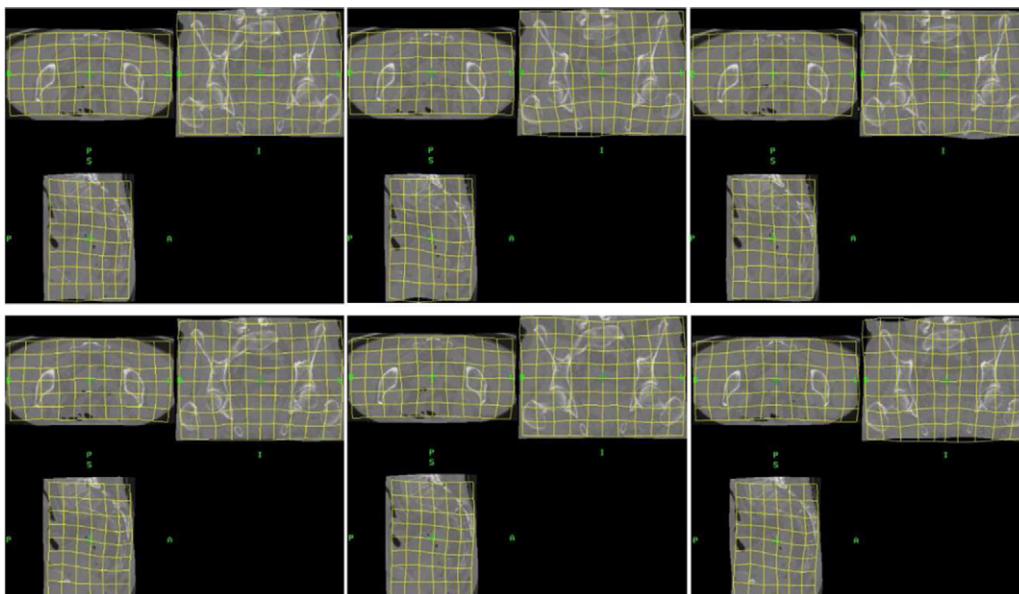


Fig. 6. Leading 3 models of variation from mean in SDM.

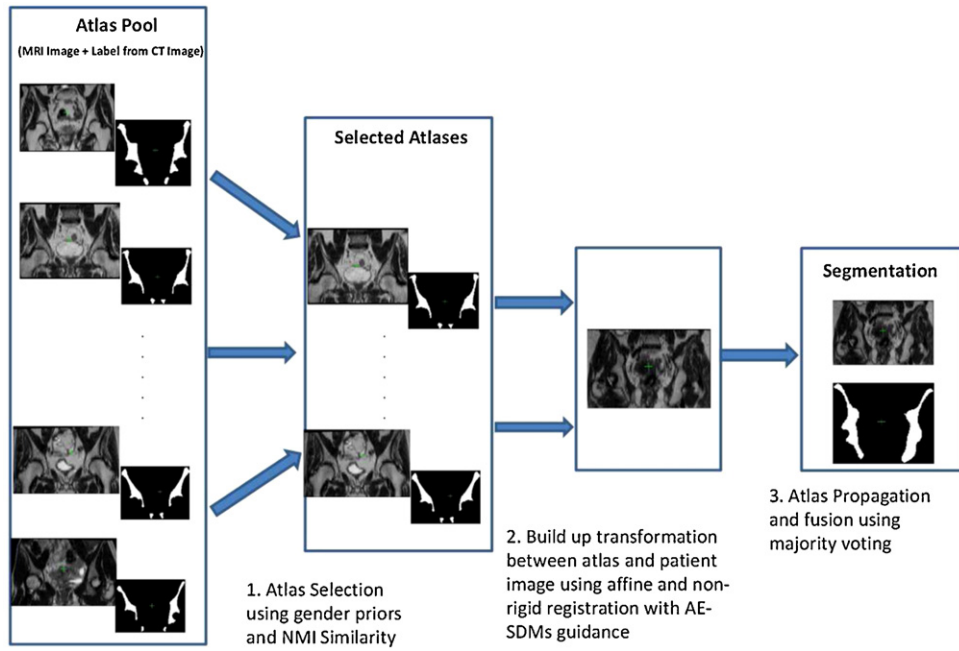


Fig. 7. The pipeline of multi-atlas segmentation of bony pelvis.

male $\Phi(M)$, female $\Phi(F)$ and uncertain $\Phi(U)$, which could be either gender.

$$\text{Gender}(P) = \begin{cases} \text{Male}, & \text{if } \theta_p < 82 \\ \text{Uncertain}, & \text{if } 82 \leq \theta_p \\ \text{Female}, & \text{if } \theta_p \geq 90 \end{cases} \quad (5)$$

Fig. 8 demonstrates male and female sub-pubic angle difference, and the coarse gender classification result from the given training sample.

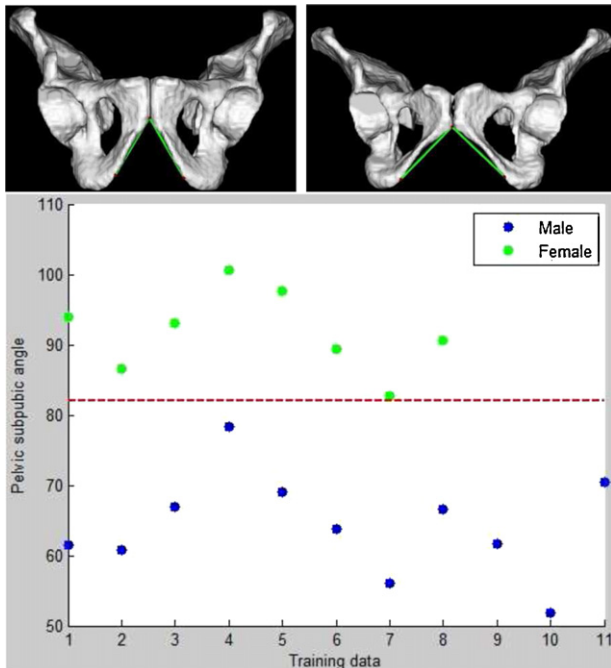


Fig. 8. The sub-pubic angle difference between male (top left) and female (top right). The graph shows the pubic angles of our training sample.

We sort the sub-pubic angle of the atlas first, and select the top M atlases whose sub-pubic angle is closest to the target. Then, to optimize atlas selection further, the value of intensity-based normalized mutual information (NMI) after registration is introduced to choose more similar atlases according to this intensity information. In the first step, nonrigid transformations between each atlas and the template image have been established. Then the N atlases (usually 10–12) with the highest NMI between the deformed atlases and the deformed target are selected for multi-atlas segmentation.

$$\text{NMI}(A, B) = \frac{H(A) + H(B)}{H(A, B)} \quad (6)$$

where $H(A)$, $H(B)$ denote the marginal entropies of A , B and $H(A, B)$ denotes their joint entropy, which is calculated from the joint histogram of A and B .

4.2. Atlas to MRI target registration using AE-SDMs guidance

Direct alignment of the bone structure from our training atlas MRI datasets to a new MRI has proven unreliable in experiments. In order to tackle this problem, we introduce an SDM-guided nonrigid registration to improve the initial alignment. The optimization path is restricted along the subspace defined by the SDM modes, which is considered to be the subspace of feasible transformation. The new transformation can be denoted by the following equation:

$$T(X) = T_{\text{global}}(X) + T_{\text{local}}(X) = (AX + t) + T_{\text{local}}^{\text{AE-SDM}}(X) \quad (7)$$

where the global transformation is obtained from affine registration and the local transformation $T_{\text{local}}^{\text{AE-SDM}}(X)$ is initialised as the mean deformation field in the AE-SDM. Gradient projection is used to restrict gradient descent to the SDM space, which means no flexibility is allowed for irregular deformation, especially pathological deformation.

However, as only 30 training data sets are used to capture the deformation field in this paper, the SDM may not cover the whole feasible variation space. Therefore, directly applying SDM in the whole non-rigid registration process may produce a worse result in some cases with unusual shape variations. In this paper, we propose

Table 1

Average metric compared with golden standard, including: volumetric overlapping accuracy (OA), dice metric (DM), average symmetric surface distance (ASD), average symmetric roots mean square surface distance (ASRMS).

Segmentation method	OA (average +SD) [%]	DM (average +SD)	ASD (average +SD) [mm]	ASRMS (average +SD) [mm]
Single-atlas (MRI)	80.1259 ± 4.0088	0.8315 ± 0.0352	2.2355 ± 0.3849	3.0112 ± 0.5514
Multi-atlas (10 MRI atlas)	86.9489 ± 4.0046	0.8930 ± 0.0341	1.3977 ± 0.4095	1.8829 ± 0.5818
Multi-atlas (10 CT atlas)	75.1439 ± 5.3561	0.8086 ± 0.0341	2.5587 ± 0.2989	3.2088 ± 0.3941
AE-SDM based multi-atlas	88.3552 ± 3.1472	0.9058 ± 0.0234	1.2377 ± 0.2723	1.5968 ± 0.3686

an initial non-rigid registration with 30 mm control point spacing using statistical deformation priors captured by the AE-SDMs. The outcome will be to warp the MRI atlas to closely represent a likely target shape. Then, a two-level non-rigid registration without the AE-SDM constraint is introduced to further improve the registration results, with the control point spacings set to 20 mm and 10 mm, respectively. The results show that our method improves the registration result, especially in some difficult cases. The registration between MRI atlases and diagnostic resolution MRI targets is concluded as follows:

Step 1: Initial global transformation using rigid transformation, followed by 9-DOF affine registration.

Step 2: One-level non-rigid registration is used to capture the coarse local transformation using AE-SDM constraints during gradient projection with 30 mm control point spacing.

Step 3: Two-level non-rigid free-form registration is introduced to further improve the registration results, with the control point spacing set to 20 mm and 10 mm, respectively.

4.3. Multi-atlas fusion

During multi-atlas fusion process, it is reasonable to expect the atlases whose reference images are more similar to the target image will contribute more from those reference images are less similar to the target image. To combine the labels for multi-atlas segmentation, we introduce similarity-ranking based weighting strategy for multi-atlas fusion [29], described as

$$\Gamma_i = \underset{s \in \{1 \dots S\}}{\operatorname{argmax}} \sum_j I(D_{ij} = s) \times w_j \quad (8)$$

where Γ_i is the estimated label for voxel i , s is the segmented labels, D_{ij} is the segmentation decision of atlas j , $I(\cdot)$ is the label function and w_j is the weight of atlas j . Here we define the weight w_j as the normalized mutual information (NMI) between registered atlas and target image:

$$w_j = \operatorname{NMI}(T_{ij}, I_{\text{Template}}) \quad (9)$$

5. Dense visual tracking using the preoperative model

The segmented preoperative model can be used as a prior map for dense visual tracking, provided the model is manually registered as described earlier and projectively textured with the surgical scene.

5.1. Perspective camera model

A 2D point set \mathbf{p} can be projectively projected from a 3D surface model using a pinhole camera model:

$$\mathbf{p} = \pi(KT\mathbf{x}) \quad (10)$$

where \mathbf{x} is a homogeneous 3D point set on the surface model and K is a 3×4 calibrated camera intrinsic matrix. The point set \mathbf{x} in a world coordinate system is first transformed to a local camera coordinate

system using a rigid body transformation $T \in \mathbb{SE}(3)$ which is a 4×4 homogeneous matrix:

$$T = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \quad (11)$$

where $R \in \mathbb{SO}(3)$ is a 3×3 rotation matrix and $t \in \mathbb{R}^3$ is a 3-vector translation. The transformed point set is then projectively projected by left-multiplying K and de-homogenized using the function $\pi: \mathbb{R}^3 \rightarrow \mathbb{R}^2$.

5.2. Dense visual tracking

Dense camera tracking can be achieved by estimating a live camera pose T_{live} frame by frame. Specifically, for dense visual tracking using a photometric cost-function, the 3D points on the model surface are projected into a live image to find their corresponding color or intensity and a residual with a reference image is calculated. It follows that the reference image \mathbf{I}_{ref} is obtained by projecting the textured surface model using the previous tracked pose T_{ref} and is registered with the warped image formed using the evaluated pose T_{live} to map the 3D surface points onto the incoming live video image \mathbf{I}_{live} . The photometric error F can thus be defined as:

$$F(\delta) = \frac{1}{2} \sum_{\mathbf{p} \in \Omega} (f_{\mathbf{p}}(\delta))^2, \quad (12)$$

$$f_{\mathbf{p}}(\delta) = \mathbf{I}_{\text{live}}(\pi(KT_{\text{live}}(\delta)\mathbf{x})) - \mathbf{I}_{\text{ref}}(\mathbf{p}) \quad (13)$$

where Ω is the set of projection points of the visible surface. The advantage of having a textured model in our framework is that the reference image \mathbf{I}_{ref} and the warped live image \mathbf{I}_{live} can be obtained by directly projecting the surface points.

Eq. (12) is zero when the projection image \mathbf{I}_{live} with respect to the pose $T_{\text{live}}(\delta)$ is equivalent to the reference image \mathbf{I}_{ref} . To evaluate the live camera pose, we use *Lie algebra* to minimally parametrize $\delta = \log_{\mathfrak{se}(3)}(T_{\text{live}})$, and $\delta \in \mathfrak{se}(3)$ is in the tangent space of the manifold $\mathbb{SE}(3)$ around the identity $\mathbf{0}$ [30]. Starting at $T_{\text{live}} = T_{\text{ref}}$, in each iteration $T_{\text{live}}(\delta) = \exp_{\mathbb{SE}(3)}(\delta)$ is updated as $\delta \leftarrow \delta + \partial\delta$ where $\partial\delta$ is evaluated by solving the stationary point of the derivative function $F(\delta) = 0$. Eq. (12) can be minimized using Gauss–Newton, Levenberg–Marquardt [31] or efficient second order (ESM) algorithms [32].

5.3. Motion blur and robust estimation

To deal with the motion blur caused by rapid camera motion, a standard image pyramid is build. The coarse-to-fine scheme can catch large motion while providing a full resolution image in the lowest layer to refine the tracking. Furthermore, optimizing in the higher layers of the pyramid image assists the tracker in avoiding local minima [5].

Since a dense method takes the entire image into account, any observed point that does not belong to the surface model may have a great impact on the tracking. In addition, other outliers can be potentially caused by specular highlights, occlusion and deformation. A robust estimation is therefore essential in our application.

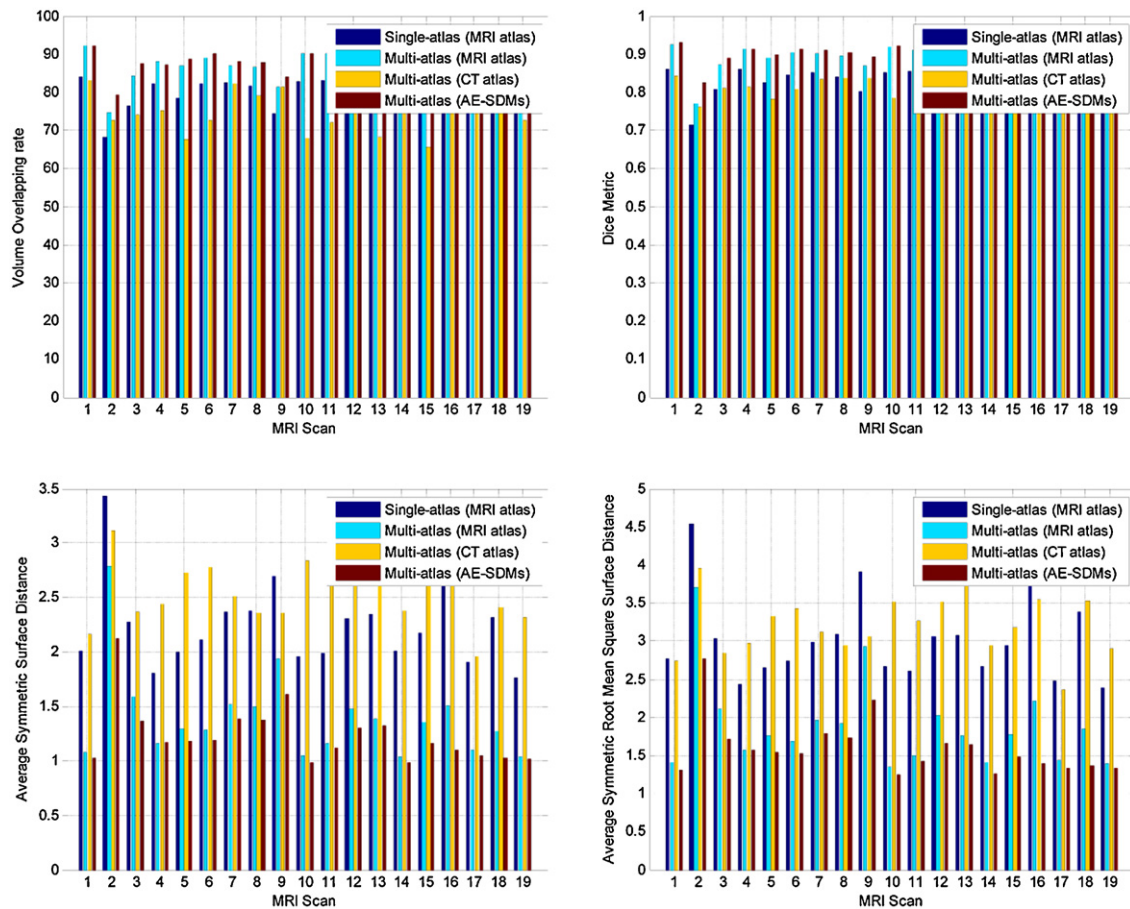


Fig. 9. Evaluation results of stated algorithm (AE-SDMs guided multi-atlas segmentation) compared with single atlas segmentation, direct multi-atlas segmentation using CT as atlas and direct multi-atlas segmentation using MRI as atlas with different metric: OA (top left), DM (top right), ASD (bottom left) and ASRMS (bottom right).

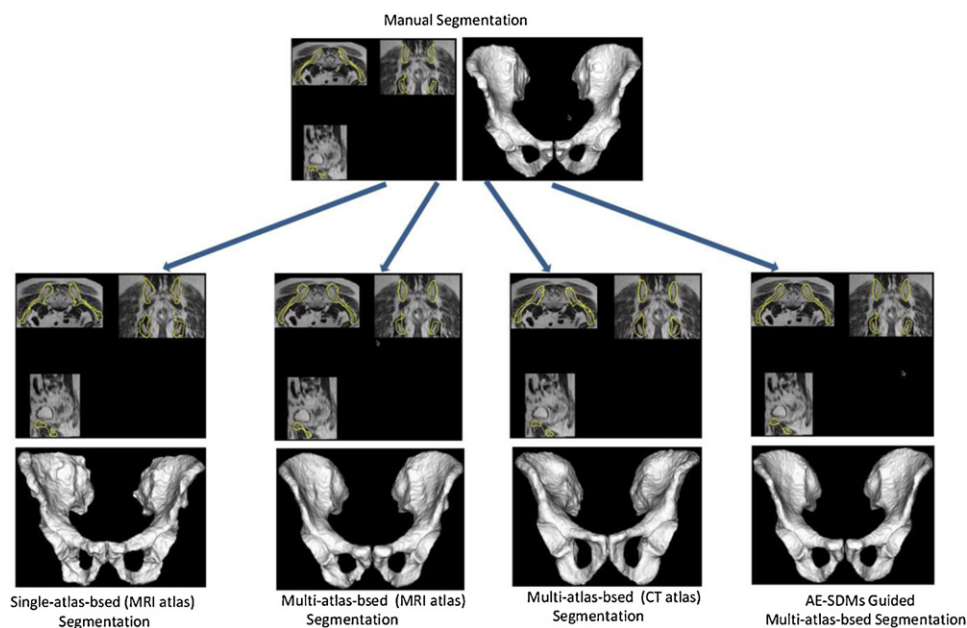


Fig. 10. MRI bony pelvis segmentation, from left to right: single atlas based, multi-atlas based (MRI), multi-atlas based (CT), multi-atlas based with AE-SDMs guidance.

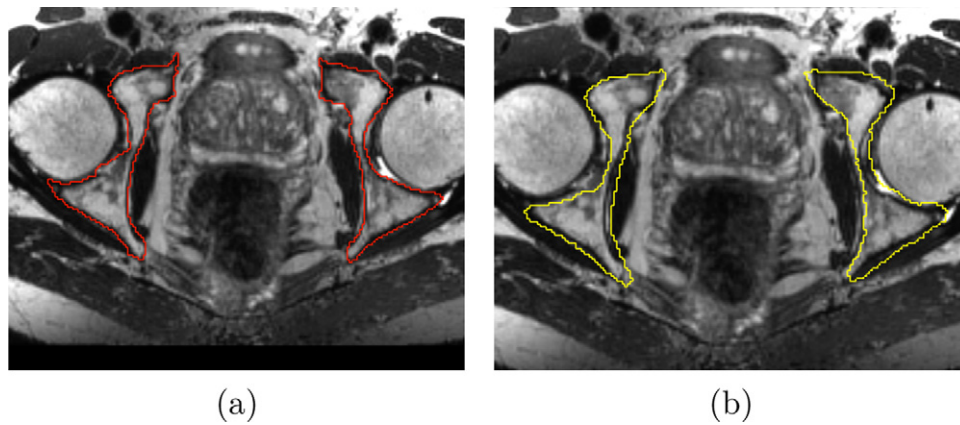


Fig. 11. Manual (a) and automatic (b) MRI segmentation results for our RALP patient example.

We remove the outliers by applying an estimator function to Eq. (14).

$$F_{robust}(\delta) = \frac{1}{2} \sum_{p \in \Omega} \rho(f_p(\delta))^2 \quad (14)$$

Here we use a simple Tukey estimator which has an asymptotic nature and treats all points greater than a certain distance as outliers [33].

6. Experiments and results

The whole system is implemented in C++ and CUDA and the evaluation is run on quad 3.20 GHz CPUs and a graphic card with 96 CUDA cores and 1 GB of global memory.

6.1. Bony pelvis segmentation

We have 30 anonymised CT scans which cover the full bony pelvis. Of those, 19 CT scans have both their corresponding diagnostic MRI image and manually segmented labels which serve as gold standards and multi-atlas labels. We carried out a set of leave-one-out studies among all these 19 MRI scans for a robust bony pelvis segmentation applying the algorithm presented in Section 3. The most similar 10 atlases are selected for bony pelvis fusion. For each data set to be segmented, each SDM was generated excluding the respective pelvis from the training sets, which ensures the independence of training and testing data. For evaluation, we used the following metric compared with the gold standard:

(1) Volumetric overlapping accuracy (OA):

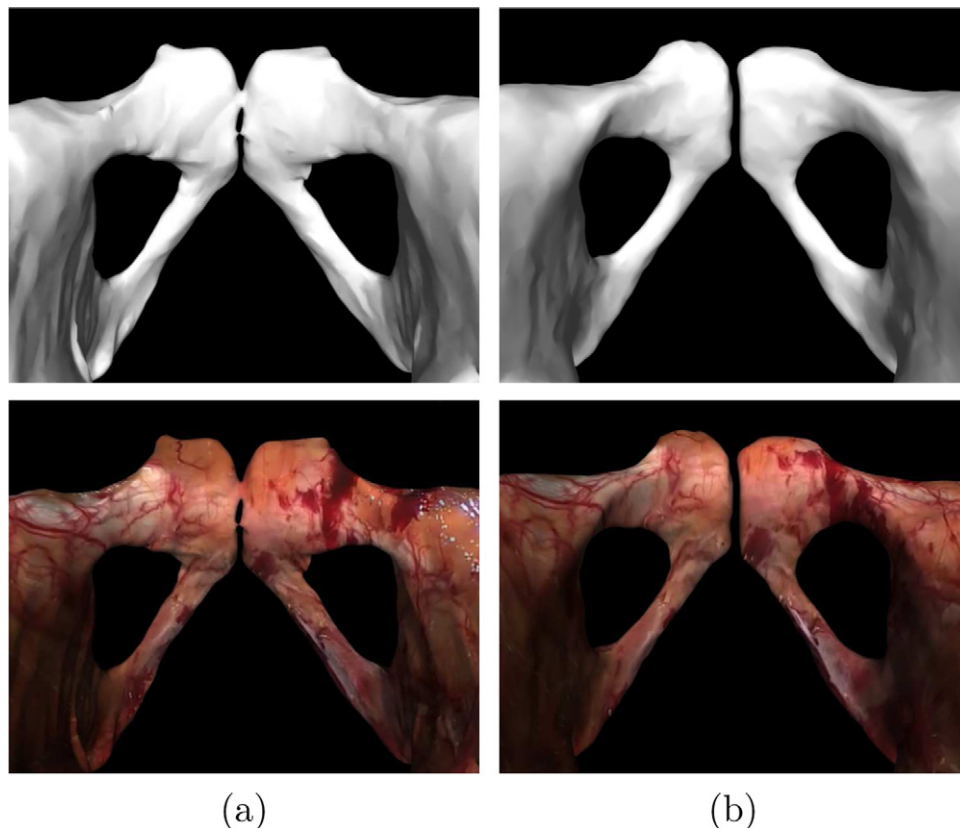


Fig. 12. Gold standard segmented model and its texture appearance (a). The model segmented by the proposed approach and its texture appearance (b).

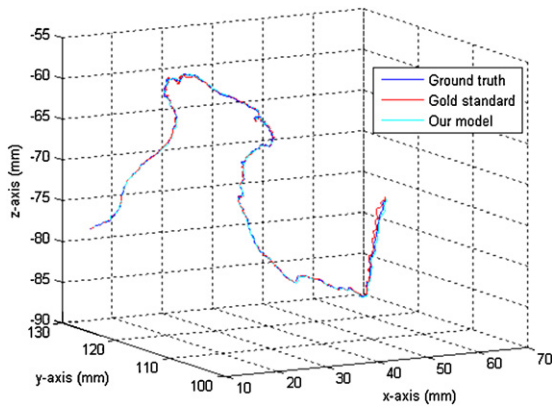


Fig. 13. Tracking trajectories in 700 frames. The ground truth trajectory is in blue. Tracking using a gold standard model is in red. Tracking using the model segmented by the proposed segmentation approach is in cyan. (For interpretation of color in the artwork, the reader is referred to the web version of the article.)

$$OA = \min \left(\frac{A \cap B}{A} \times 100\%, \frac{A \cap B}{B} \times 100\% \right).$$

(2) Dice metric (DM):

$$DM = 2 \frac{A \cap B}{A + B}.$$

(3) Average symmetric surface distance (ASD):

$$ASD = \frac{\sum_{i=1}^N |(S_{A_i} - T(S_{B_i}))| + \sum_{i=1}^N |(S_{B_i} - T^{-1}(S_{A_i}))|}{2N}.$$

(4) Average symmetric roots mean square surface distance (ASRMS):

$$ASRMS = \sqrt{\frac{\sum_{i=1}^N (S_{A_i} - T(S_{B_i}))^2 + \sum_{i=1}^N (S_{B_i} - T^{-1}(S_{A_i}))^2}{2N}}.$$

Table 1 shows the average value and standard deviations of these metrics among all the 19 MRI scan segmentations compared with manual segmentation from CT scans. Final segmentation results for the complete bony pelvis from MRI images exhibit an average OA of $88.3552 \pm 3.1472\%$, an average DM of 0.9058 ± 0.0234 , an average ASD of 1.2377 ± 0.2723 mm and an average ASRMS of 1.5968 ± 0.3686 mm.

The segmentation result for each of the MRI scans is shown in Fig. 9. From these comparisons, we can see that multi-atlas using the original MRI as the atlas produced a better result than a single atlas segmentation. However, multi-atlas segmentation with CT as the atlas and AE-SDM-guided multi-atlas based segmentation further improved the accuracy, especially for some poor alignments, which is clearly shown in Fig. 10.

Table 2

The average position errors over 700 frames and the average running times in each tracking frame.

	Average position error (mm)	Average running time (ms)
Gold standard model	0.16	251
Our model	0.13	253

6.2. Intraoperative tracking

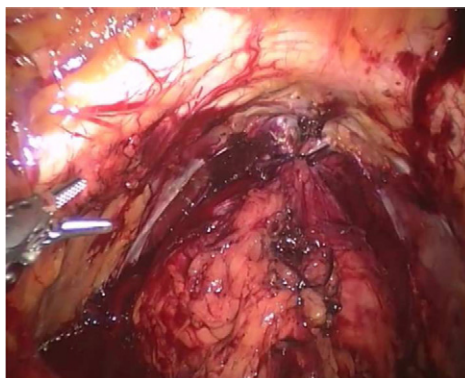
We applied our automated segmentation algorithm to a particular clinical case, where the patient then went on to have a robotic prostatectomy. In the experiment, the model manually segmented by an expert is regarded as the gold standard. The segmentation results can be seen in Fig. 11. We then manually align the two models with the real surgical scene to texture them as shown in Fig. 12. This manual alignment should be as geometrically accurate as possible since we need a correct pose to projectively texture the surface model. The surface textured model will be used in subsequent dense tracking.

We first track the camera in a clinical video with the gold standard model to obtain a camera trajectory and regard this as a ground truth. To evaluate the feasibility of using the proposed automatic segmentation with the proposed dense tracking, we construct a simulation platform and use the ground truth trajectory to imitate the camera motion. The residual of the camera positions between the ground truth trajectory and the evaluated trajectory is of particular interest.

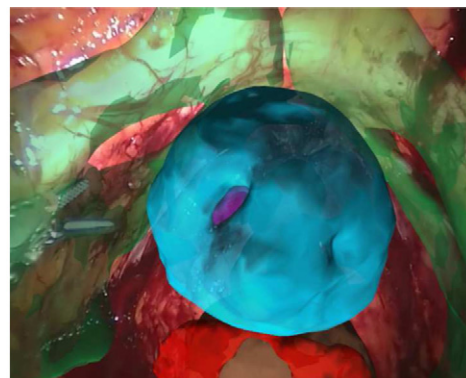
Fig. 13 shows the tracking trajectories of 700 frames in the simulation environment, and the corresponding average position errors and average running time are shown in Table 2. The trajectory which is very close to the ground truth means the tracking is robust, and we can consistently augment virtual models into the scene for image-guided surgery and the corresponding video as shown in Fig. 14. Indeed, perhaps due to its smoothness, using the automated segmentation slightly outperformed using the manual segmentation (0.13 mm compared to 0.16 mm average position error). Note that both of the measured errors are very small is because the simulation environment is assumed to be a Lambertian reflectance and no deformation in the scene, which also leads highly robust tracking in both cases.

7. Discussion and future work

We have shown how the bony pelvis can be accurately and robustly segmented using a combination of statistical deformation



(a)



(b)

Fig. 14. An augmented reality example. The original video scene (a). The corresponding augmented scene (b). This is best illustrated in the accompanying video (Appendix A).

modeling and multi-atlas techniques. An MRI appearance embedding statistical deformation model (AE-SDM) is proposed within a multi-atlas framework. This outperforms existing methods and provides a segmentation of the pelvis with an accuracy of 1.24 mm. The subsequent surface can then be manually registered to the surgical scene through the da Vinci stereo-endoscope [22]. Tracking is then achieved using dense correspondence via photometry. A rate of 5 frames-per-second is achievable on modest hardware, meaning that fully real-time video-based tracking should be possible. The tracking accuracy in simulation tests is seen to be 0.13 mm over 700 video frames and the tracking on a real clinical endoscopic view is seen to be stable and robust.

The run-time of our multi-atlas segmentation scheme increases the number of atlases used, with greater computation time required to get the affine and non-rigid transformation from atlas space to target space. In our experiment, 10 atlases were used, which accounts for around 2.5 hours overall run-time for one segmentation on a state-of-the-art 8-core machine. A recent implementation of our non-rigid registration algorithm using graphic processing units (GPUs) make the registration 10-times faster [34]. This makes our method quite reasonable for use in the clinical environment.

There is much further work to do. The registration of the 3D model to the surgical view needs to be automated. Initial results in the AE-SDM method could be expanded to incorporate soft tissue structures. We have identified a number of critical anatomical features that would be useful for guidance, including the urethra, prostate, seminal vesicles, dorsal venous complex and neurovascular bundle [3]. These must be segmented by hand at present. Rigid registration of the pelvis is a good starting point for guidance, but there is significant deformation of soft tissue structures due to surgery, pneumoperitoneum or the inherent movement of the prostate with respiration. These could be approximated using a physical model or the likely deformations could be learned from a series of datasets.

We present the first system for automated identification of the bony pelvis from MRI and tracking of this registered model in the surgical scene in real time. This represents a significant step towards clinically useful augmented reality guided robotic prostatectomy, which has potential to improve both functional and oncological outcomes and reduce the learning curve in RALP.

Acknowledgements

This research was funded by Cancer Research UK under project A8087/C24250. The Pelican Foundation also funded parts of this research under the Pelvic Anatomy Model (PAM). The authors are grateful for support from the NIHR Biomedical Research Centre funding scheme. We are also grateful to the radiology and theatre staff at the Imperial College Healthcare NHS Trust for their help and cooperation throughout this project. The work has ethical approval from the London-Dulwich research ethics committee.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.compmedimag.2013.01.001>.

References

- [1] Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA: Cancer J Clin* 2011;61(2):69–90.
- [2] Kaul S, Saveria A, Badani K, Fumo M, Bhandari A, Menon M. Functional outcomes and oncological efficacy of Vattikuti Institute prostatectomy with veil of aphrodite nerve-sparing: an analysis of 154 consecutive patients. *BJU Int* 2006;97:467–72.
- [3] Cohen D, Mayer E, Chen D, Anstee A, Vale J, Yang GZ, et al. Augmented reality image guidance in minimally invasive prostatectomy. In: Madabhushi A, Dowling J, Yan P, Fenster A, Abolmaesumi P, Hata N, editors. *Prostate Cancer Imaging. Computer-Aided Diagnosis, Prognosis, and Intervention*; vol. 6367 of Lecture Notes in Computer Science. Springer Berlin/Heidelberg; 2010. p. 101–10.
- [4] Hoeks CMA, Scheenen TWJ, Vos PC. Prostate cancer: multiparametric MRI imaging for detection, localization, and staging. *Radiology* 2011;261(1):46–66.
- [5] Newcombe RA, Lovegrove SJ, Davison AJ. DTAM: dense tracking and mapping in real-time. In: 2011 international conference on computer vision 2011. p. 2320–7, doi:10.1109/ICCV.2011.6126513.
- [6] Lorigo LM, Faugeras O, Grimson WEL, Keriven R. Segmentation of bone in clinical knee MRI using texture-based geodesic active contours. In: *Medical image computing and computer-assisted intervention: MICCAI 1998*. 1998. p. 1195–204.
- [7] Rifa H, Bloch I, Hutchinson S, Wiat J, Garnerio L. Segmentation of the skull in MRI volumes using deformable model and taking the partial volume effect into account. *Med Image Anal* 2000;4(3):219–33.
- [8] Schmid J, Magnenat-Thalmann N. MRI bone segmentation using deformable models and shape priors. In: *MICCAI (1)*; vol. 5241 of Lecture notes in computer science. 2008. p. 119–26.
- [9] Schmid J, Kim J, Magnenat-Thalmann N. Robust statistical shape models for MRI bone segmentation in presence of small field of view. *Med Image Anal* 2011;15(1):155–68.
- [10] Lamecker H, Seebass M, Hege HC, Deuffhard P. A 3D statistical shape model of the pelvic bone for segmentation. *Proc SPIE: Med Imaging* 2004;5370(section 4):1341–51.
- [11] Seim H, Kainmueller D, Heller M, Lamecker H, Zachow S. Automatic segmentation of the pelvic bones from CT data based on a statistical shape model. In: *Eurographics workshop on visual computing for biomedical (2008)*. 2008. p. 93–100.
- [12] Vasilache S, Ward K, Cockrell C, Ha J, Najarian K. Unified wavelet and Gaussian filtering for segmentation of CT images; application in segmentation of bone in pelvic CT images. *BMC Med Inform Decision Making* 2009;9(Suppl 1):S8, doi:10.1186/1472-6947-9-S1-S8.
- [13] Thompson S, Penney G, Buie D, Dasgupta P, Hawkes D. Use of a CT statistical deformation model for multi-modal pelvic bone segmentation. *Proc SPIE Med Imaging* 2008.
- [14] Gooya A, Pohl KM, Bilello M, Cirillo L, Biros G, Melhem ER, et al. GLISTR: glioma image segmentation and registration. *IEEE Trans Med Imaging* 2012;31(10):1941–54, doi:10.1109/TMI.2012.2210558.
- [15] Lötjönen JM, Wolz R, Koikkalainen JR, Thurfjell L, Waldemar G, Soininen H, et al. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *NeuroImage* 2010;49(3):2352–65.
- [16] Babalola KO, Patenaude B, Aljabar P, Schnabel J, Kennedy D, Crum W, et al. Comparison and evaluation of segmentation techniques for subcortical structures in brain MRI. In: *MICCAI (1)*; vol. 5241 of Lecture notes in computer science. 2008. p. 409–16.
- [17] Mei L, Figl M, Rueckert D, Darzi A, Edwards PJ. Sample sufficiency and number of modes to retain in statistical shape modelling. In: *MICCAI (1)*; vol. 5241 of Lecture notes in computer science. 2008. p. 425–33.
- [18] Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging* 1999;18(8):712–21, doi:10.1109/42.796284.
- [19] Frangi A, Rueckert D, Schnabel J, Niessen W. Automatic construction of multiple-object three-dimensional statistical shape models: application to cardiac modeling. *IEEE Trans Med Imaging* 2002;21(9):1151–66.
- [20] Rueckert D, Frangi AF, Schnabel JA. Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *IEEE Trans Med Imaging* 2003;22(8):1014–25, doi:10.1109/TMI.2003.815865.
- [21] Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *NeuroImage* 2009;46(3):726–38.
- [22] Pratt P, Mayer E, Vale J, Cohen D, Edwards E, Darzi A, et al. An effective visualisation and registration system for image-guided robotic partial nephrectomy. *J Robot Surg* 2012;6(1):23–31.
- [23] Chang PL, Chen D, Cohen D, Edwards PE. 2D/3D registration of a pre-operative model with endoscopic video using colour-consistency. In: *Proceedings of the 6th MICCAI workshop on augmented environments for computer-assisted interventions*; vol. 7264 of Springer LNCS. 2011. p. 1–12.
- [24] Schnabel JA, Rueckert D, Quist M, Blackall JM, Castellano-Smith AD, Hartkens T, et al. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In: *MICCAI*; vol. 2208 of Lecture notes in computer science. 2001. p. 573–81.
- [25] Commowick O, Malandain G. Efficient selection of the most similar image in a database for critical structures segmentation. In: *MICCAI*; vol. 4792 of Lecture notes in computer science. 2007. p. 10–203.
- [26] Wu M, Rosano C, Lopez-Garcia P, Carter CS, Aizenstein HJ. Optimum template selection for atlas-based segmentation. *NeuroImage* 2007;34(4):1612–8.
- [27] Igbigbi PS, Nanono-Igbigbi AM. Determination of sex and race from the subpubic angle in Ugandan subjects. *Am J For Med Pathol* 2003;24(2):168–72.

- [28] Bertino AJ. Forensic science – fundamentals and investigations. South-Western Cengage Learn 2000:368.
- [29] Wang H, Suh JW, Das S, Pluta J, Altinay M, Yushkevich P. Regression-based label fusion for multi-atlas segmentation. In: CVPR 2011. 2011. p. 1113–20, doi:10.1109/CVPR.2011.5995382.
- [30] Gallier J. Geometric Methods and Applications: For Computer Science and Engineering. Texts in Applied Mathematics. Springer; 2001.
- [31] Baker S, Matthews I. Lucas-kanade 20 years on: a unifying framework. Part 1. Tech. Rep. CMU-RI-TR-02-16. Robotics Institute; 2002.
- [32] Malis E. Improving vision-based control using efficient second-order minimization techniques. In: Proceedings of IEEE international conference on robotics and automation (ICRA); vol. 2. 2004. p. 1843–8.
- [33] Comport AI, Marchand E, Chaumette F. Statistically robust 2-D visual servoing. IEEE Trans Robot 2006;22(2):415–20.
- [34] Modat M, Ridgway GR, Taylor ZA, Lehmann M, Barnes J, Hawkes DJ, et al. Fast free-form deformation using graphics processing units. Comput Meth Programs Biomed 2010;98(3):278–84, doi:10.1016/j.cmpb.2009.09.002.