Automatic Inference and Measurement of 3D Carpal Bone Kinematics from Single View Fluoroscopic Sequences

Xin Chen, Jim Graham, Charles Hutchinson, and Lindsay Muir

***Abstract*—We present a novel framework for estimating the 3D poses and shapes of the carpal bones from single view fluoroscopic sequences. A hybrid statistical model representing both the pose and shape variation of the carpal bones is built, based on a number of 3D CT data sets obtained from different subjects at different poses. Given a fluoroscopic sequence, the wrist pose, carpal bone pose and bone shapes are estimated iteratively by matching the statistical model with the 2D images. A specially designed cost function enables smoothed parameter estimation across frames and constrains local bone pose with a penalty term. We have evaluated the proposed method on both simulated data and real fluoroscopic sequences and demonstrated that the relative poses of carpal bones can be accurately estimated. One condition that may be assessed using this measurement is dissociation, where the distance between the bones is larger than normal. Scaphoid-Lunate dissociation is one of the most common of these. The error of the measured 3D Scaphoid-Lunate distances were 0.75 $\pm$ 0.50 mm for simulated data (25 subjects) and 0.93 $\pm$ 0.47 mm for real data (15 subjects). We also propose a method for constructing a 'standard' pathology measurement tool for automatically detecting Scaphoid-Lunate dissociation conditions, based on single-view fluoroscopic sequences. For the simulated data, it produced 100% sensitivity and specificity. For the real data, it achieved 83% sensitivity and 78% specificity.**

***Index Terms*—Carpal bone poses, 2D 3D registration, Statistical pose model, Statistical shape model, Fluoroscopic sequence, Wrist pathology**

I. Introduction

WRIST pain, either acute or chronic, is a common presenting symptom in hand clinics. It may be due to a number of different pathologies, including acute trauma, arthritis (either osteo or inflammatory), vascular disorders, the sequelae of congenital abnormalities and the sequelae of trauma. These latter may include osteoarthritis secondary to fracture malunion or nonunion, and ligament instability. The standard assessment of a patient with pain of this nature will include history taking, clinical examination and special investigations.

X. Chen and J. Graham are with the Centre for Imaging Science, Faculty of Medicine, Manchester Academic Health Sciences Centre, The University of Manchester, U.K. (e-mail: xin.chen@manchester.ac.uk)

C. Hutchinson is with the Division of Healthsciences, University of Warwick, U.K.

L. Muir is with the Department of Hand Surgery, Salford Royal NHS Foundation Trust, Salford, U.K.

This work was supported by Medical Research Council, U.K., under Grant 87997.

The wrist joint is complex, and the maintenance of the normal relationship of the carpal bones, both at rest and on movement is governed by intercarpal and extrinsic ligaments. Normal function and integrity of these ligaments is essential for the smooth movement of the wrist. No tendons insert onto the carpal bones themselves, and their movements are therefore dictated by the movements of the surrounding bones. Ligamentous injuries may lead to disordered movements of the bones. These disordered movements in turn lead to abnormal loading and hence to osteoarthritis. Standard assessment of these disordered movements includes plain radiography, MR scanning and cine radiography. The first two modalities give static images that may readily be examined and measurements taken, but are only static images of a dynamic problem. Cine radiography (e.g. fluoroscopic sequences) is more subjective and requires judgement and experience. If there is still doubt about the diagnosis, wrist arthroscopy may give further information, but this is an invasive procedure and therefore entails risk and expense. A method of determining carpal kinematics from fluoroscopic sequences that allowed more objective evaluation would be of value to the hand surgeon in accurate diagnosis. It would also contribute to treatment evaluation and to understanding an area of hand surgery that still remains challenging.

Here we present a method for computer interpretation of the fluoroscopic sequences to attain a higher degree of objectivity and quantification in the diagnostic process. The wrist is a complex joint (see Fig. 2); during wrist movement, the eight carpal bones follow a complex, multi-dimensional trajectory, making interpretation of radiographs difficult. One important step towards this aim is the development of statistical models (SM) of the carpal bones and their spatial relationships during movement, which is able to represent the pose and bone shape variation in much fewer dimensions. For this study we have trained this SM from a set of CT images from different subjects at different poses. Subsequently, the full 3D carpal bone motions can be recovered by matching the SM with the fluoroscopic sequences through 2D-3D image registration techniques.

A number of studies have sought to represent the carpal kinematics using CT or MR data, mainly concentrating on representing 'average' kinematics over a small number of individuals (e.g. [1], [2]). Van de Giessen et al. [3] presented a 3D rigid registration method based on segmented meshes, which aims to build a SM of carpal bones. More recently, they introduced a 4D statistical model that locally describes

the relative positions of the carpal bones [4] in pre-defined poses, with the aim of detecting abnormal bone spaces. A comparison of wrist poses captured statically and dynamically was reported in [5]. They concluded that negligible differences were observed between the dynamic motion and the step-wise static motion of the carpal bones from 'healthy' subjects. Some authors have focussed on building hierarchical statistical shape models ([6], [7]) or an articulated shape model [8]. Davatzikos et al. [6] presented a method of using the wavelet transform to capture different levels of shape detail in a coarse to fine structure, which enables the statistical shape model to cover a larger range of variability with a small number of training samples. Cerrolaza et al. [7] further extended the idea to deal with multiple objects for 2D brain image segmentation, where the objects to be included for model building at each level have to be carefully selected. Boisvert et al. [8] studied spine variation using 3D articulated pose models. The relative rigid transformation parameters of each vertebra with respect to the vertebra of the upper level were used to construct the articulated pose model. The spine variations between the same set of patients before and after treatment were compared using the model. Point-based statistical models, such as [7] do not retain the rigidity of each of the multiple objects. In our proposed framework we build a statistical pose model (SPM), based on geometrical transformation parameters and a separate point-based statistical shape model (SSM) to deal with the issue of shape variation and articulation of the carpal bones. We use the combined model to fit to image sequences for quantifying 3D movement.

Many studies have investigated 2D-3D image registration (e.g. [9], [10], [11]), mainly in the field of registration of pre-operative MR or CT images to intra-operative 2D images. Our work differs from these in that we seek to achieve registration of a 2D image sequence to a 3D model (not derived from the same individual) to infer the poses and shapes of an individual wrist. Zheng [12] took a similar approach to estimate the orientation of the pelvis from a single X-ray image. Whitmarsh et al. [13] presented a method to reconstruct both the 3D bone shape and 3D areal bone mineral density distribution of the proximal femur from a single dual-energy X-ray absorptiometry image. More recently, Baka et al. [14] and Zheng et al. [15] similarly presented a statistical shape model based framework to estimate femur shapes from multiple X-ray images. In the case of [14], fluoroscopic sequences were used, similarly to the work reported here.

The main contributions of this paper, distinguishing it from these earlier studies, are: (1) A hybrid SM is developed representing both the complex pose and shape variation of the eight carpal bones plus radius and ulna. (2) The full 3D motion and bone shapes are recovered by matching the SM with a single view fluoroscopic sequence: a difficult ill-posed problem. (3) Our initial results show that the relative positions between the carpal bones can be estimated accurately through the proposed framework. (4) We have constructed a pathology detection tool that takes advantage of the inherent ability of the SPM to align wrist poses. In [4], they also detect abnormal bone spaces based on 3D input data sets for limited number of pre-defined flexion-extension poses. We are not aware of any study which attempts to make a 2D to 3D inference and measurement in a system of this level of complexity. An early version of this work was published in [16]. In this paper we describe the framework in greater detail and report the following further developments: (1) The SPM presented here is generated based on both the radial-ulnar poses and flexion-extension poses, where the SPM used in [16] is only based on radial-ulnar poses. (2) Faster optimisation and more robust registration, arising from the use of a more constrained model. (3) Additional registration accuracy is achieved by the use of local pose refinement, controlled by a new cost function term. (4) Rather than building a SSM for each individual bone, all bones are modelled simultaneously to represent the shape variations of the ensemble of bones. This helps to maintain the nature of the relationships between adjacent bone shapes and reduces the number of shape parameters. (5) We include more comprehensive experimental results based on real fluoroscopic sequences using extended training datasets. (6) A method of constructing the pathology detection tool, based on the SPM, is introduced for the first time. The evaluation results demonstrate the feasibility of using the proposed system for clinical diagnosis.

The overview of the proposed framework is illustrated in figure 1(a). The system consists of a training phase and a 2D-3D image registration phase. We currently have CT data from 25 subjects, each at five poses (neutral pose and two extreme poses in flexion-extension and radial-ulnar deviation). The segmentation of each bone and rigid registration parameters that align bones at different poses within and across the subjects in the training set were obtained using an iterative segmentation and registration algorithm [17]. Segmentation results were confirmed by an experienced radiologist. A hybrid statistical model, representing both the pose and shape variations, was built from the results of the segmentation-registration framework. The SPM was built based on the transformation parameters, while the segmentation result was used to build the SSM. In the 2D-3D image registration phase, the global 3D rigid transformation, the poses of carpal bones, the local 3D rigid transformation of each bone and the bone shapes were estimated iteratively in sequence from each frame of the fluoroscopic video. The registration is performed sequentially, frame-by-frame, the estimated poses at each frame acting as the starting positions for the next (see figure 1(a)). Detailed descriptions are given in the following sections.

II. Coordinate System and Problem Parameterisation

In figure 1(b), the coordinate axes X^S, Y^S and Z^S define the source coordinate system with the origin at the radiation source, whereas X^M, Y^M and Z^M define the machine coordinate system with the origin at the isocentre. u and v define the image coordinates, normal to the direction of the radiation beam. The origin of the image plane is at the projection of the optical centre.

In order to interpret the true 3D motion of the carpal bones, four sets of parameters are estimated iteratively in sequence during image registration: (1) Rigid transformation

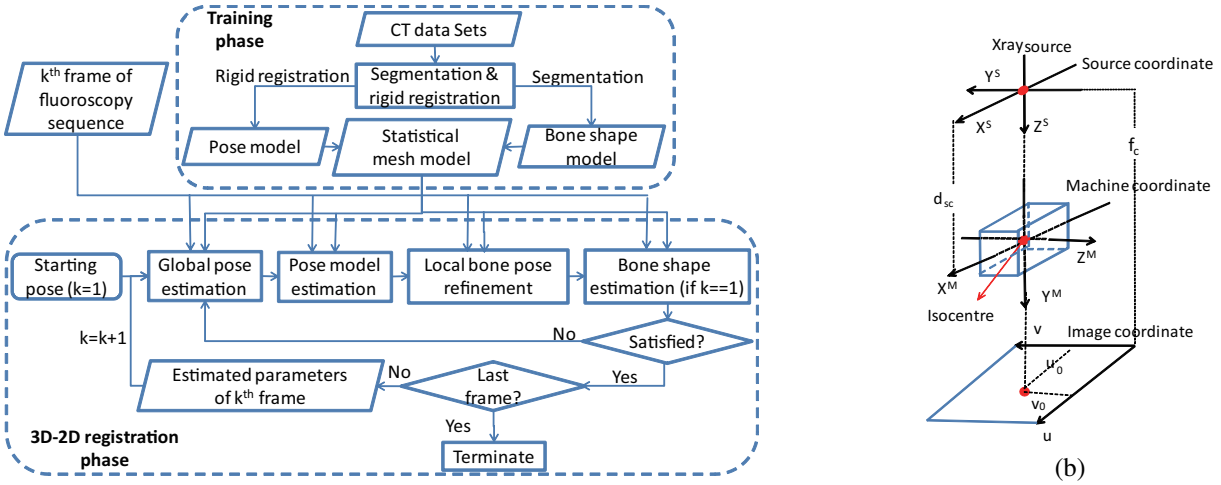

Fig. 1. (a) Overview of the proposed system. (b) Perspective projection geometry for the fluoroscopic imaging system.

parameters of the wrist and a global scale factor, denoted by θ=$\{tx,\ ty,\ tz,\ r1,\ r2,\ r3,\ s\}$ in the machine coordinates. t=$[tx, ty, tz]^T$ denotes the translations along X^M, Y^M and Z^M axes. r=$[r1, r2, r3]^T$ is the set of Rodrigues parameters [18] representing the global orientations. The magnitude of vector r is the rotation angle around the axis represented by the normalised unit vector of r. s controls the distance between the centroid of each bone and the origin in the radius, and the global size of the bones. (2) SPM parameters b^m represent the carpal bone poses during movement. By using the pose model parameters, the transformation parameters of each bone can be obtained, denoted as m_i=(tx_i^m, ty_i^m, tz_i^m, $r1_i^m$, $r2_i^m$, $r3_i^m$) (i is an index identifying each bone) (3) Transformation parameters l_i=(tx_i^l, ty_i^l, tz_i^l, $r1_i^l$, $r2_i^l$, $r3_i^l$, s_i^l) of each bone used to refine the poses estimated from the pose model. (4) SSM parameters b^q for bone shape estimation. Using homogenous coordinates, the constructed 3D statistical mesh model can be projected to the image plane by,

$$A_i = KTPD_i \begin{bmatrix} ss_i^l Q_i \\ 1 \end{bmatrix} \tag{1}$$

where Q_i indicates the mesh points of the estimated shape for the i^{th} bone. s and s_i^l are the global and local scale factors respectively that control the size of the carpal bones. D_i is the pose matrix of the i^{th} bone estimated using the pose model and the local pose refinement. P is the global rigid transformation matrix. T is the transformation matrix from the machine coordinate system to the source coordinate system, and K is the intrinsic projection matrix of the X-ray imaging system. In detail, P is denoted as:

$$P = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \tag{2}$$

where t is the translation vector $[tx, ty, tz]^T$. R is the 3×3 rotation matrix represented by Rodrigues parameters [4], [18], calculated as

$$R = I + B\sin|r| + B^2(1 - \cos|r|) \tag{3}$$

where $|r|$ is the magnitude of the orientation vector [r1,r2,r3]^T. I is the identity matrix, and B is the skew-symmetric matrix normalised by $|r|$, expressed as

$$B = \begin{bmatrix} 0 & r3 & -r2 \\ -r3 & 0 & r1 \\ r2 & -r1 & 0 \end{bmatrix} /|r| \tag{4}$$

In equation (1), D_i is calculated as,

$$D_i = \begin{bmatrix} R_i^g & t_i^g \\ 0 & 1 \end{bmatrix} \tag{5}$$

where t_i^g=$s[tx_i^m, ty_i^m, tz_i^m]^T$+$[tx_i^l, ty_i^l, tz_i^l]^T$ is the summation of translation vectors estimated from the pose model and local bone refinement. $R_i^g = R_i^m R_i^l$ is the 3× 3 rotation matrix that combines the rotations estimated from pose model and local bone refinement respectively. R_i^m and R_i^l can be calculated individually by equation (3) using their corresponding Rodrigues parameters.

Furthermore, T in equation (1) is given by

$$T = \begin{bmatrix} C_{az} & S_{az} & 0 & 0 \\ -S_{el}S_{az} & S_{el}C_{az} & C_{el} & 0 \\ C_{el}S_{az} & -C_{el}C_{az} & S_{el} & d_{SC} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6}$$

where C and S denote cosine and sine functions, subscripts az and el denote the view angles for 3D-2D projection, with $az = 0°$ and $el = 180°$ producing the anteroposterior view, $az = 90°$ and $el = 180°$ producing the left lateral view, and $az = -90°$ and $el = 180°$ producing the right lateral view. d_{SC} indicates the distance between the isocentre and the X-ray source.

K in equation (1) is given by

$$K = \begin{bmatrix} \frac{f_c}{pix_u} & 0 & u_0 & 0 \\ 0 & \frac{f_c}{pix_v} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{7}$$

where f_c is the distance between the X-ray source and detector plane. pix_u and pix_v are the physical pixel sizes along the horizontal and vertical directions of the detector, and (u_0, v_0) are the coordinates of the optical centre on the image. In our data, (u_0, v_0) were always the centre of the image. pix_u and

pix_v are known from the detector specification. Therefore, only f_c needs to be estimated, which can be done by measuring the distance between the X-ray source and detector for each subject.

The use of Rodrigues parameters to represent bone orientations is convenient for pose model building and parameter optimisation. More importantly, unlike the quaternion representation, it does not require vector normalisation, nor does it suffer from the singularity problem that arises when using the Euler angle rotations.

III. Training of Pose Model and Shape Model

To generate training data it was necessary to achieve consistent segmentations and poses of the bones across subjects in the training set and across the five wrist positions within each subject. For this we developed an integrated framework [17] that combines the Grow Cut segmentation method with rigid image registration to simultaneously segment and align the carpal bones in the CT data sets. The kinematics of the carpal bones is complex and significant pose differences can be introduced as the joint adopts different positions. The framework significantly reduces the workload of segmentation, while simultaneously providing a good alignment of the carpal bones. Each bone segmentation was verified by an experienced radiologist.

As the shape of each bone may vary from individual to individual, we modelled this variation using a point distribution model (PDM) [19]. This was built using the segmented volume of the same set of training subjects. Correspondence between these surfaces of bones across subjects was established by the minimum description length (MDL) algorithm [19]. The 3D structure of each bone is described by a set of 1002 points on the segmented surface. In our earlier work [16] we modelled the shape of each bone independently. However here we maintain the nature of the relationships between adjacent bone shapes and reduce the number of shape parameters by representing the shape points of all bones in a single column vector in a consistent order. One training example is described by (x_1, y_1, z_1, ...x_{10020}, y_{10020}, z_{10020}) ^{t} (10 bones $\times$ 1002 points each). The coordinates of the shape points of each bone are expressed with respect to its own centroid, eliminating any linkage between the shape model and the pose model. The deformable shape model is then described as,

$$q = \mu^q + \phi^q b^q \tag{8}$$

where μ^q and ϕ^q (superscript q is a notation indicating the shape parameters) are the mean shape and the principal subspace matrix for the shapes. b^q is the shape model parameter to be estimated. We retain the first 15 significant components in the shape model, which keeps about 90% of variation.

The statistical pose model was trained using the six rigid transformation parameters. The common coordinate system for all pose parameters has an origin at a specified point in the radius for a reference subject. The sizes of all the wrists are normalised to a consistent scale. The pose of one subject is described by (tx_1, ty_1, tz_1, $r1_1$, $r2_1$, $r3_1$, ..., tx_{10}, ty_{10}, tz_{10}, $r1_{10}$, $r2_{10}$, $r3_{10}$)^t. (8 carpal bones, 1 radius and 1 ulna). The orientation parameters allow for a continuous description of the wrist movement (see section II). Then the pose model can be parameterised as,

$$m = \mu^m + \phi^m b^m \tag{9}$$

where the mean pose μ^m (superscript m is a notation indicating the pose model parameters) and the principal subspace matrix ϕ^m are computed from 5 (poses)$\times$ 25 (subjects) training samples using PCA. The vector b^m represents the pose parameters that describe the pose (m) along each principal direction. In our experiments, only the first two significant components are used, which keeps 90% of variation. The first component reflects the flexion-extension motion and the second component represents the radial-ulnar motion. By contrast, our earlier work [16] used 8 significant modes representing 98% of the variation based on 10 training subjects for the radial-ulnar movement only. Experimentally we found that the use of fewer model components reduced computational time by 40%. The inclusion of flexion-extension poses for training also extends the motion range which helps to reduce the registration errors. This probably arises because, in capturing the training data, there was no constraint on the radial-ulnar movement in CT, so that the correspondence between the extreme positions of radial-ulnar movement in CT and fluoroscopy may not be exact. There are also potentially small differences in the directions of flexion-extension and radial-ulnar movement between the fluoroscopy and CT image capture processes. By further combining with the local bone refinement procedure, the more constrained model achieved smaller registration error in 3D by around 0.7mm (values shown in table I), compared with the results in [16]. Our experiments also showed that including significant components beyond two does not improve the registration accuracy, which indicates that the local bone refinement process (section IV) dealt with the deviation from the linear pose model very well.

Based on the SSM and the SPM, a hybrid statistical mesh model can be built by using the Crust mesh construction algorithm [20]. Figure 2 shows the poses of the first two components of the SPM (represented by the mean shapes of each bone) and the first mode of the shape variation.

IV. 3D-2D Image Registration

The statistical mesh model is then used to match with each of the frames in the fluoroscopic sequence to infer the 3D motion and bone shapes (see figure 1(a)). The position of the model is firstly initialised interactively by indicating a central point on the radius in the first frame of the fluoroscopic sequence. Then the poses of the bones in each frame are estimated in sequence, the poses from the current frame being used as the starting poses of the next. Figure 3 summarises the registration process, in which the preprocessed fluoroscopic image is iteratively matched with a simulated projection generated from an updated pose of the mesh model. This registration procedure is used specifically in the pose estimation and refinement steps illustrated in figure 1(a). For each iteration, the global pose parameter θ=$\{tx, ty, tz, r1, r2, r3, s\}$, the SPM parameter b^m, the local transformation

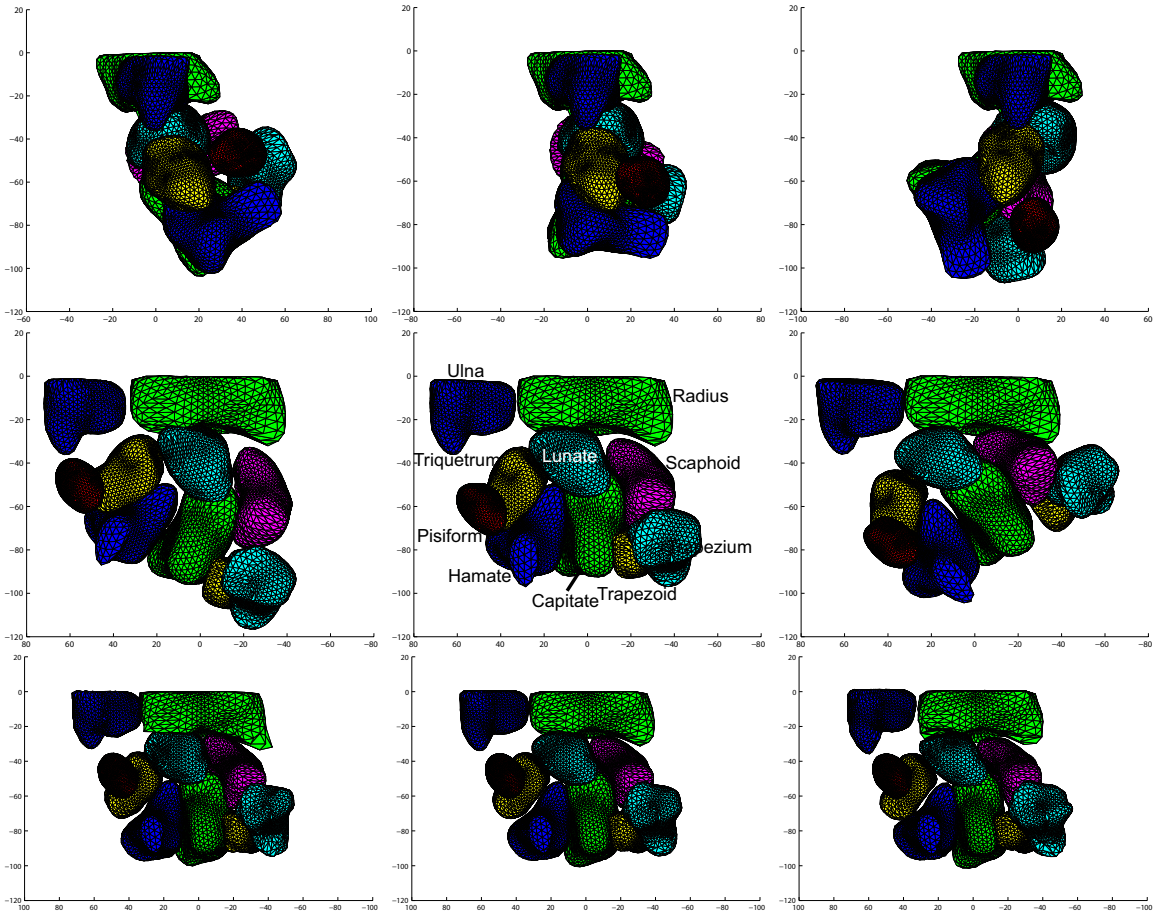

Fig. 2. Top row: The poses of the first component of the pose model (lateral view) that mainly describes the flexion-extension movement. Middle row: The poses of the second component of the pose model (AP view) that mainly represents the radial-ulnar movement. Bottom row: the first component of the shape model. (Major shape variations occur in the Ulna, Radius and Lunate.) In each case the mean $\pm$ 2 s.d. are shown.

parameters l_i=(tx_i^l, ty_i^l, tz_i^l, $r1_i^l$, $r2_i^l$, $r3_i^l$, s_i^l) of each bone, and the SSM parameters b^q are updated iteratively in sequence. Detailed descriptions are given in the following subsections.

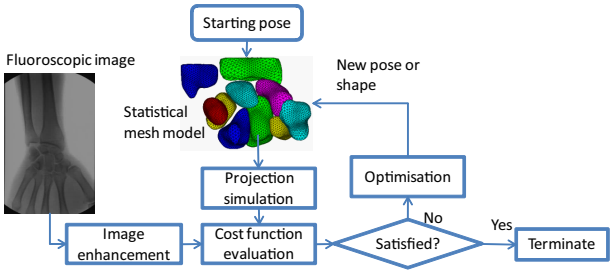

Fig. 3. Overview of the 3D-2D image registration process.

A. Fluoroscopic image enhancement and projection simulation

As there is considerable variation in the quality of fluoroscopic images, preprocessing is necessary to achieve consistent results. Firstly the intensities are normalised to zero mean and unit standard deviation. This is followed by anisotropic diffusion [21] to smooth the image and preserve edges. Local gradients are used for image matching, and figure 4 shows the gradient maps of an original fluoroscopic image, and the image after anisotropic diffusion and after normalisation and anisotropic diffusion respectively.

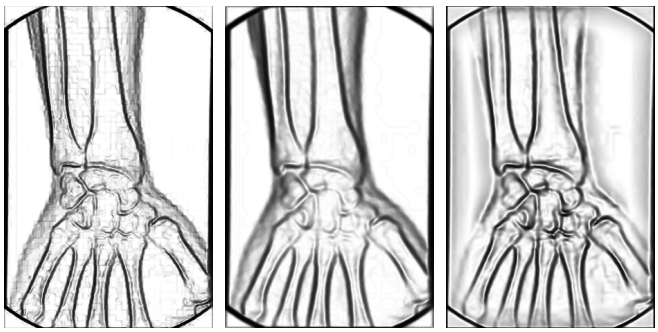

Fig. 4. Left: the gradient map of the original flouroscopic image. Middle: the gradient map of the image after applying the diffusion filter. Right: the gradient map of the image after local normalisation and diffusion.

In order to optimise the pose parameters, we iteratively generate projections from the statistical mesh model with updated pose parameters. The mesh model is considered as a binary volume with background set to zero and bone set to unity. Based on the perspective projection model described in section II, the simulated projection can be generated by ray

casting. The projected intensity is in negative proportion to the sum of binary values along the ray from the source to each pixel in the image plane. The simulated image that represents the mean shape and mean pose of the model is shown in figure 5.

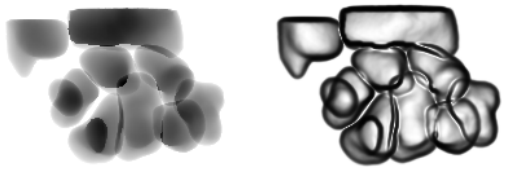

Fig. 5. Left: simulated image that represents the mean shape and mean pose of the model. Right: the magnitude of gradient.

B. Cost function

To evaluate the similarity between the fluoroscopic image and the simulated image, we investigated several forms of the cost function, achieving best results from the one shown in equation (10), based on the gradient along horizontal and vertical directions as well as the gradient magnitude of the two images. Additionally, the adjacent frames to the current fluoroscopic image were also taken into account in the cost function to make the estimated poses smooth across frames.

If we define the Normalised Correlation Coefficient (NCC) between two images A and B as C(A,B), then the proposed cost function can be described as:

$$E_1 = -C(Om_{k-1}, Om_k) - \sum_{p=k-1,k,k+1} w_p(C(Im_p, Dm_k) + C(Ix_p, Dx_k) + C(Iy_p, Dy_k)) \quad (10)$$

where k is the current frame number of the fluoroscopic sequence. Im_p, Ix_p and Iy_p are the gradient magnitude, vertical gradient and horizontal gradient of the fluoroscopic image at the p^{th} frame respectively. Dm_k, Dx_k and Dy_k are the corresponding values of the simulated image. The use of the absolute gradient magnitude in the second term, in addition to the signed gradient, results in a smoother objective function, resulting in a reduced tendency to converge to local minima than is the case when using signed gradients alone. Calculating the cross-correlation between sets of three adjacent frames makes the estimated pose smooth across frames. The inter-frame weighting parameters, w_{k-1}, w_k and w_{k+1} were set at 0.2, 0.6 and 0.2 respectively. For the first term of the cost function, the vertices in the statistical mesh model are projected to the image plane; we assume the intensities at those projected points are similar across adjacent frames. Om_{k-1} and Om_k represent the gradient magnitude of the previous frame and the current frame at the projected correspondence positions. The first term makes the shape of the cost function sharper, leading to a faster and more accurate optimisation result. The $(k-1)^{th}$ frame and $(k+1)^{th}$ frame are not evaluated for the first and last frame respectively.

Equation (10) is used to estimate the global pose parameter θ and the SPM parameter b^m. The wrist motion can be described as approximately linear by the SPM parameters, where the deviations from linear positioning are accommodated by the local refinement of individual bone poses. In the local refinement procedure, a different cost function is used, where an additional term is added to E_1 as described in equation (11). The additional term makes the estimated local pose as close as possible to the pose model, weighted by a Gaussian distribution. This is able to preserve the topology of the carpal bones, when the intensity term E_1 is weak.

$$E = E_1 + \omega \exp(-\frac{\frac{1}{p}\sum_{i=1}^{p} ||x_i^g - T^l(x_i^g)||^2}{2\phi^2}) \quad (11)$$

In equation (11), x_i^g represents the i^{th} 3D mesh points after the global pose and pose model estimation. p is the total number of mesh points of the currently evaluated bone (In our case, p=1002 for each bone). T^l is the local transformation matrix for that bone. ω is the weighting parameter that balances the image intensity term E_1 and the added geometric penalty term. ϕ is the standard deviation of the Gaussian distribution. In our evaluation tests, $\omega = -0.2$ and $\phi = 10$ were experimentally determined and used.

C. Optimisation

The coordinate origin for all motions is the centroid of the radius in the statistical mesh model. The global transformation parameters are estimated based on the regions surrounding all bones and iteratively refined by alternating with the SPM parameter and local transformation parameter estimations (see figure 1(a)). By estimating the SPM parameter b^m based on all carpal bone and ulna regions, a set of transformation parameters $m_i = \{tx_i^m, ty_i^m, tz_i^m, r1_i^m, r2_i^m, r3_i^m\}$ that representing the kinematic pose of the i^{th} bone can be generated by equation (9). The local transformation parameters l_i for each bone are calculated by evaluating the cost functions on the corresponding bone volumes. The set of mesh points Q_i that represent the i^{th} bone shape are obtained by substituting the estimated SSM parameter b^q into equation (8). Subsequently, the 2D projection that represents the current estimated 3D pose of the carpal bones can be generated using equation (1).

The optimisation method we have used is a simplified version of the Brent-Powell method [22], requiring a smaller number of optimisation steps. We used parabola fitting to replace the Brent line search in the Brent-Powell method. The multi-dimensional search space (θ, b^m, l_i and b^q) is explored by iterative individual 1D line searches. For each parameter search, the cost function is evaluated three times at the current position and its negative and positive neighbours, respectively, with the initial distance between the current position and its neighbours pre-defined by a search range. To fit a parabola to these three values, the following three criteria are applied to select the best parameter value for the next iteration:

(1) A minimum is found by equating the first derivative of the fitted parabola to zero, with the second derivative being positive: In this case, the minimum is selected as the current

best parameter value for the next iteration or for evaluation of the next parameter.

(2) A maximum is found with the second derivative being negative: In this case, the parameter value corresponding to the smallest cost function value, evaluated at the current position and its neighbours, is selected.

(3) A minimum is found, but it is too far away from the evaluated position (located outside twice the initial search range due to the cost function being too flat): In this case, the transformation parameter value corresponding to the smallest cost function value, evaluated at the current position and its neighbours, is again selected.

When a better value is found for one parameter, it will be used for evaluating the next one. When all the transformation parameters satisfy the first criterion, the search range is reduced by dividing it by a factor to refine the estimation results. The whole optimization is terminated when the changes in the evaluated cost function values are smaller than a pre-set threshold or the search range is small enough.

In our case, the true sizes of the bones are unknown; recovering the 3D pose from a single image is therefore a difficult, ill posed, problem. Any movement along the Y^M axis in the machine coordinates, could be compensated by scaling of the bone. In order to minimise this effect, the optimisation is carefully sequenced. We firstly assume that the wrist is not moving along Y^M axis during radial-ulnar movement (ty=0), as it is placed on a flat surface. Following the interactive initialisation (section IV), the first frame is evaluated, taking all the parameters into account (except ty) in the following sequence: tx, tz, $r1$, $r2$, $r3$, s, b^m, tx_i^l, ty_i^l, tz_i^l, $r1_i^l$, $r2_i^l$, $r3_i^l$, s_i^l and b^q. After convergence, the estimated pose of the current frame is used as the starting pose for the next frame. The global scale factor s, local scale factors s_i^l and shape model parameters b^q are only estimated once in the first frame. From our initial experiments, the shape parameters are not improved significantly when we include more frames and the fitting is made significantly more complex and time consuming.

V. Evaluation

The true 3D poses corresponding to the recovered poses for real fluoroscopic sequences are not available: there is no ground-truth against which to judge the accuracy of the recovered poses. This would require the synchronisation of 3D imaging with the fluoroscopic imaging devices. The proposed framework was therefore evaluated based on 25 simulated sequences in addition to 15 real fluoroscopic sequences. All evaluations were conducted using a leave-one-out strategy, based on the training data. In all of the evaluation tests, the input fluoroscopic sequences were pre-processed to construct a 3-level multi-scale pyramid (down-sampled by a factor of 2 at each level). In the optimisation procedure, the same set of fixed initial search ranges was used at each level (4 voxels for translation, $4\pi/180$ for rotation, 0.2 for scale, one standard deviation for pose model parameters and shape parameters). The search ranges were divided by 2 each time the criteria were met (see section IV-C), and the whole process was terminated when the maximum value of the search ranges was smaller than a pre-set threshold. The registration accuracy of the simulated data and real data are shown in the following subsections. More importantly, a measurement model that represents the healthy pose of carpal bones at each kinematic pose is generated. This model can be used for pathology detection and quantification.

A. Evaluation based on simulated data

We evaluated our framework quantitatively based on a number of simulated fluoroscopic sequences generated from the 3D CT data. All CT volumes have been re-sampled to an isocubic volume with voxel dimension of 0.5 mm. We interpolated (cubic spline) a number of poses between the neutral pose and two extreme poses of radial-ulnar deviation in a particular movement cycle (neutral - full radial - neutral - full ulnar), resulting in 39 poses for each of 25 subjects. While we assume a linear model for variation in pose, the cubic spline interpolation makes the trajectory smooth around the observed poses. The ray-casting method was then used to generate a simulated fluoroscopic sequence from those interpolated 3D data. The tested dataset was not included in the training datasets.

We conducted initial leave-one-out experiments to evaluate the number of PCA components required for the SSM. In these we altered the number of shape model components, leaving all other parameters unchanged. The final 2D-3D registration accuracy stopped improving when 15 components were selected. This may be due to the shape errors estimated using components greater than 15 being less significant than the pose errors. This suggests that using 25 subjects and 15 significant components are sufficient for this application.

To test the registration accuracy of the whole framework, the 3D pose of the simulated test subject was then calculated as described in section IV. The registration error measured by the 3D Euclidean distance at each corresponding point of the mesh between the target pose and the estimated pose is presented in table I. The average 3D registration error is 2.45 $\pm$1.07 mm. The main contribution to this error is the ill-posed problem (confusion between the scale and translation along Y^M), whereas the errors along the in-plane directions, X^M and Z^M, are very small with average error about 1 voxel (0.5 mm).

As described in previous sections, the local scale factor of each bone is also estimated in the local refinement procedure to take account of the fact that the relative sizes of bones will vary between individuals. Based on the 25 independent tests, the mean value of this local scale factor varies between 0.98 and 1.13, depending on which bone is being considered. The standard deviations are around 0.05, indicating that the relative sizes of bones varies between individuals. While this complicates the optimisation, the last column of table I shows that the optimisation without the local scale results in a larger registration error.

In clinical diagnosis, the absolute positions of the carpal bones in 3D space are not important; of greater significance is the relative movement of the bones. By using our method, the relative positions of the carpal bones with respect to

TABLE I
THE AVERAGE ERROR IN MM, MEASURED IN 3D, X^M, Y^M AND Z^M AXES, BETWEEN THE TARGET AND ESTIMATED CORRESPONDENCE POINTS OF EACH CARPAL BONE OF 25 SUBJECTS: TRIQUETRUM(TRI), LUNATE(LUN), SCAPHOID(SCA), PISIFORM(PIS), HAMATE(HAM), CAPITATE (CAP), TRAPEZOID (TRD) TRAPEZIUM (TRM). THE "NO LOCAL SCALE" COLUMN LISTS THE REGISTRATION ERRORS WITHOUT ESTIMATION OF LOCAL SCALE OF EACH BONE (SEE TEXT).

	eTri	eLun	eSca	ePis	eHam	eCap	eTrd	eTrm	Average	No local scale
Err3D	2.51±1.42	2.23±1.24	2.18±1.50	2.63±1.72	2.31±1.44	2.34±1.50	2.51±1.72	2.86±1.85	2.45±1.07	2.86±1.08
ErrX	0.59±0.44	0.58±0.50	0.44±0.37	0.58±0.53	0.44±0.33	0.45±0.35	0.53±0.44	0.47±0.43	0.51±0.39	0.60±0.41
ErrY	2.15±1.56	1.38±1.35	1.43±1.60	2.24±1.84	2.10±1.55	2.11±1.61	2.23±1.82	2.64±1.94	2.16±1.14	2.53±1.20
ErrZ	0.68±0.55	0.59±0.49	0.48±0.39	0.68±0.61	0.44±0.35	0.45±0.35	0.55±0.46	0.57±0.52	0.55±0.43	0.64±0.46

each other can be estimated much more accurately than the absolute positions of the individual bones. In calculating the distance between bones we use the average distance between corresponding surface points. Each bone is represented by the same number of surface points (1002), determined when the shape model was constructed using the MDL method (section III). Correspondences are determined using the index of each point, giving a consistent set of correspondences.

One condition that may be assessed using this measurement is dissociation, where the 3D distance between the bones is larger than normal. As an example of this, we investigate Scapholunate dissociation, which is one of the most common of these conditions. The registration error of the 3D distance between the Lunate and the Scaphoid (dLS) was measured. The error is 0.75 ± 0.50 mm, compared to an average surface to surface distance of 2 mm between the Scaphoid and Lunate. The surface to surface distance is measured by the average of the 20 shortest Euclidean distances between the surface points of the two bones. More importantly, using the statistical model, the measured 3D bone distances can be normalised to a consistent scale by dividing them by the estimated global scale factor s and an average of the two bones' local scale factors, calculated as $(s_i^l + s_j^l)/2$. This leads to automatic classification of the bone dissociation cases, which could not possibly achieved without the statistical model (section V-C).

B. Evaluation based on real data

We also tested our framework on 15 real fluoroscopic sequences. There were about 40-100 frames per sequence, covering the radial-ulnar movement. In the absence of ground truth, the absolute positions of bones cannot be used for evaluation. However, the key question is whether the estimated relative distances between bones are equivalent to the measurements from CT data, and the diagnostic conclusions unchanged. The registration accuracy of the real data can be validated by comparing the 3D distance between Scaphoid and Lunate (dSL) estimated from real fluoroscopic sequences and the original 3D volumes of the same subject.

One major advantage of using the SPM as one of the registration steps is that the kinematic pose of the wrist from different motion sequences can be aligned directly based on the SPM parameters. This provides an advantage compared with the method described in [4] where it is required to align the wrist to pre-defined discrete poses. The first two components of our SPM cover 90% of variation in the full range of flexion-extension and radial-ulnar movements. The combination of the two components is also able to generate interpolated poses within the motion range. To measure the error in the estimated 3D pose at each wrist position, we need to compare it with the pose of the 3D CT data at that position. To do this, we need to index the positions along the motion trajectory, which can be done using the first two components of the SPM. The values of these components define corresponding positions for the model and the CT data.

To produce ground truth corresponding to the original 3D data requires 3D-3D registration between each bone in the 3D statistical mesh model and the corresponding bone in the original 3D data set. This was done at a number of poses by estimating the global pose parameter, SPM parameter, and local pose parameter (θ, b^m, l_i-section IV) at each pose. The CT volume was then set to the same pose location according to the first two components of the estimated SPM parameter. Poses of the original 3D data, other than the neutral and extreme poses were generated by cubic spline interpolation. Having matched the poses of the estimated and real 3D bone positions, the 3D distances between the Scaphoid and Lunate in the original and estimated volumes were measured and normalised using the estimated global scale factor (s). The 3D-3D registration was achieved using a method similar to that described in [17]. This is not the main focus of this paper, so we do not provide details of the implementation here.

Another important issue is the reliability of the 2D-3D registration, as it may give mis-aligned results due to low quality of the fluoroscopic sequence. Since the kinematic pose represents the 'average' pose of the carpal bones, the local deviation from the kinematic pose should be relatively consistent across the sequence. A particular frame showing a larger deviation from the kinematic pose than other frames may indicate a failed registration at that frame. Hence, the 3D Euclidean distance between the local refined bone pose and the kinematic pose is used to indicate the reliability of the registration, which is calculated by equation (12).

$$r = \frac{1}{p}\sum_{i=1}^{p} ||x_i^g - T^l(x_i^g)||^2 \tag{12}$$

In equation (12), x_i^g represents the i^{th} 3D mesh points after the global pose and SPM estimation. p is the total number of mesh points of the current evaluated bone (In our case, p=1002 for each bone). T^l is the local transformation matrix for that bone. Then the value r is subtracted from the mean deviation

TABLE II
ERRORS (MM) OF ESTIMATED 3D SCAPHOID-LUNATE DISTANCE BETWEEN REAL FLUOROSCOPIC SEQUENCES AND THE CORRESPONDING 3D VOLUMES.

Subjects	1	2	3	4	5	6	7	8
Error	0.77±0.60	1.35±0.46	0.65±0.43	0.45±0.32	0.80±0.30	0.30±0.12	1.02±0.05	1.20±0.68
Subjects	9	10	11	12	13	14	15	Average
Error	1.38±0.24	0.19±0.22	0.87±0.47	0.90±0.74	0.75±0.45	0.99±0.50	2.37 ± 1.40	0.93±0.47

$\bar{r}$ of the whole sequence. This is denoted as δr.

The registration was considered as successful if the deviation δr is smaller than 1 voxel (experimentally determined). Furthermore, if the smallest r is larger than a threshold, it indicates the registration of the whole sequence may not be accurate, which needs visual check by the user.

The dSL of 15 subjects were calculated, each based on the original 3D volume and real fluoroscopic sequence of the same subject at their corresponding poses. Figure 6 shows the estimated first two SPM parameters for each frame of the 15 real fluoroscopic sequences. As expected, the values of the 2^{nd} component, representing the major motion of the radial-ulnar fluoroscopic sequences, are distributed over the range of $\pm$1.5 standard deviation. The values of the 1^{st} component (representing flexion-extension motion) are within a range of $\pm$0.5 standard deviation, making a small contribution to minimising the out-of-plane transformation errors. Table II presents the mean and standard deviation of the absolute differences between the estimated dSL and the ground truth for each of the 15 subjects. Each estimated dSL was measured in the statistical model coordinate system by dividing each by their estimated scale factor, hence all the estimated dSL from different subjects can be compared at a consistent scale. 83.5% of the frames were considered as successful using the criterion based on Eqn.(12), and these were used to generate the measurements shown in table II. The average estimated error of successful registrations is 0.93 ± 0.47 mm, indicating good agreement of the dSL estimated from the real fluoroscopic sequences and the original 3D volume.

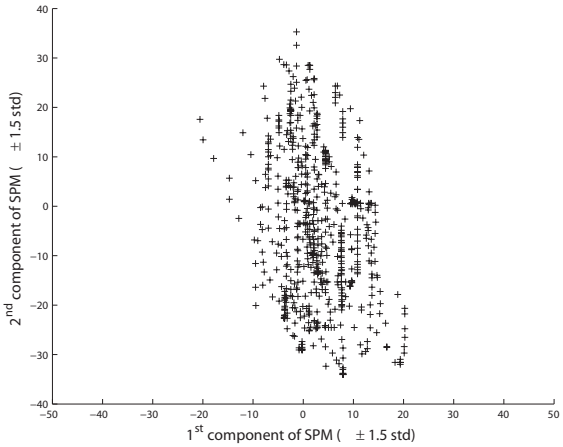

Fig. 6. Estimated first two SPM parameters for each frame of the 15 real fluoroscopic sequences.

A sample frame of the matching result and the corresponding 3D poses are shown in figure 7 in which the projected contours from the 3D mesh model are superimposed on the preprocessed fluoroscopic image. The estimated 3D mesh model in the palmar and dorsal views is shown in middle and right respectively. The registration result for the whole sequence can be found in [23].

C. Measurement Model for Pathology Detection

Our 3D CT and fluoroscopy datasets contain images of 8 and 6 individuals respectively, suffering from Scaphoid-Lunate dissociation, diagnosed radiologically on the basis of CT images. Here we demonstrate the potential to perform the diagnosis automatically from the fluoroscopic sequences.

The 3D CT volumes of 15 'healthy' subjects, assessed radiologically as not suffering from scaphoid-lunate dissociation, were used to determine a 'standard' model, based on neutral and extreme radial-ulnar poses. The statistical mesh model was aligned with these volumes by estimating the global rigid transformation parameters, the SPM parameters and the local transformation parameters for each bone (see section V-B). The kinematic poses at intermediate wrist positions were determined by cubic spline interpolation between the extreme and neutral positions, sampled at every two integer values of the second (radial-ulnar) component of the SPM, giving 36 wrist positions. In calculating the distance between bones we use the distance between corresponding surface points. As mentioned in section V-A, correspondences can be established between surface points on different bones. Here we use a reduced number of surface points (N=100, rather than 1002 used in building the model) for improved computational efficiency. Equation (13) and (14) show that we calculate the Mahalanobis distances (MD) using the means and covariances of individual pairs of corresponding points, rather than using the average distance, as in section V-A. Letting $l^k_{\phi,j}$ and $s^k_{\phi,j}$ represent the j^{th} surface point on the k^{th} sample volume at pose ϕ on the lunate and scaphoid respectively, the relative distance between the lunate and scaphoid at point j is

$$d^k_{\phi,j} = l^k_{\phi,j} - s^k_{\phi,j} \tag{13}$$

$d^k_{\phi,j}$ is a 3×1 vector, so the mean $m_{\phi,j}$ and covariance matrix $C_{\phi,j}$ of the j^{th} point pair based on all k samples at pose ϕ can be calculated. The Mahalanobis distance between the new test data and the model at pose ϕ is calculated using equation (14).

$$m_\phi = \frac{1}{N}\sum_{j=1}^{N}\sqrt{(d^{new}_{\phi,j} - m_{\phi,j})^T C^{-1}_{\phi,j}(d^{new}_{\phi,j} - m_{\phi,j})} \tag{14}$$

To assess a new wrist, the 2D radial-ulnar fluoroscopic sequence can be registered with the statistical model using

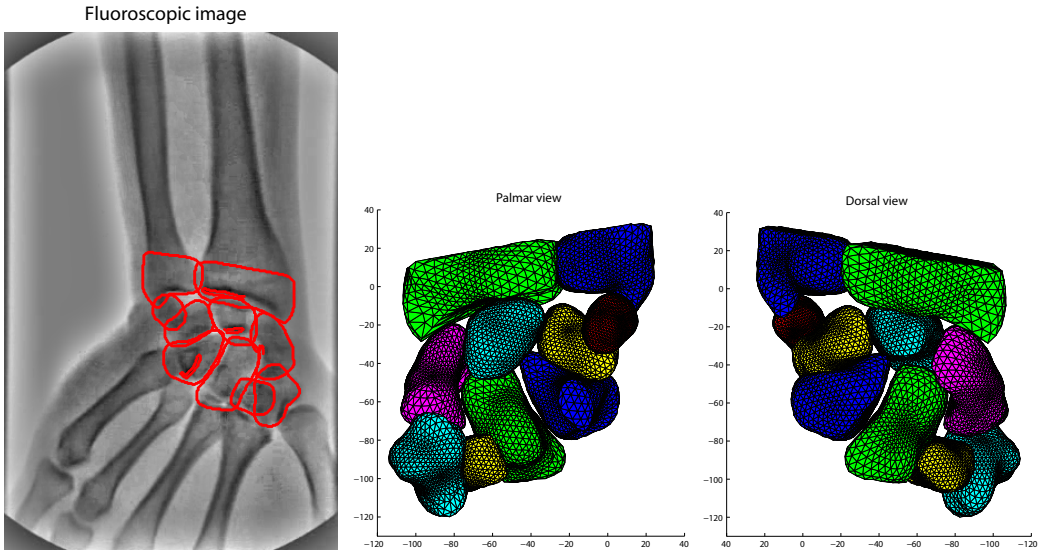

Fig. 7. Registration result of one frame from a real fluoroscopic sequence. The registration result for the whole sequence can be found in [23]

the methods described in section IV, and the wrist poses determined by the 2^{nd} SPM component. The MD can then be calculated (Equation 14) at each pose to measure the deviation from the 'standard' model. The results for the 25 (17 healthy and 8 abnormal) simulated sequences and 15 (9 healthy and 6 abnormal) real fluoroscopic sequences are shown in figure 8. In this figure the triangles represent healthy subjects and the squares represent abnormal subjects. The lengths of the bars through the data points represent the reliability of each registration, as calculated in Equation (12).

As shown in Figure 8, for the simulated data, most of the abnormal subjects (squares) have larger MDs than the normal subjects (triangles). The distinction between the two groups is less pronounced for the real fluoroscopic sequences. Additionally, the registration is less reliable compared with the simulated data, due to blurring effects generated by the wrist moving too fast.

By varying the threshold (the same threshold for all kinematic poses) of MD for classifying the normal and abnormal cases, the Receiver Operating Characteristics (ROC) curve is generated and shown in Figure 9. The ROC for both the simulated data and real data are presented, using only the successful registrations (section V-B, Eqn. (12)). This resulted in using 89.3% of the frames for the simulated sequences and 83.5% of the frames for real sequences. The thresholds that produce the best error rate for simulated and real data are 2.75 and 2.86 respectively. These values result in 87.0% true positive rate (TPR) and 14.0% false positive rate (FPR) for simulated sequences, and 70.0% TPR and 30.0% FPR for real sequences.

The diagnostic conclusion for an individual can be obtained, by combining the classification results for all of the frames of the sequence. The test set for diagnosis is small, and the result rather dependent on a judicious choice of values for the MD threshold and the method used of combining the frames. We investigated two ways of deriving the classification result based on the MDs of frames. The first method is to use the weighted sum of the MD of each frame, which results in a single MD for each test sequence. The MD of each frame was weighted according to the reliability factor. The best operating point in the ROC evaluation, by varying the 'averaged' MD threshold, is found at the threshold of 2.8 which resulted in sensitivity and specificity values of 68% and 90% respectively. For the second method, we define the normal frame ratio (NFR) as the number of successful frames classified as 'normal' divided by the total number of successful frames in the assessed fluoroscopic sequence. If the NFR is greater than a threshold, the particular subject is considered as 'healthy', otherwise is diagnosed as having Scaphoid-Lunated dissociation. Figure 10 shows the ROC curve obtained by varying the NFR, using a MD threshold of 2.5 (experimentally selected) for both the simulated and real data set. The highly quantised nature of the ROC curve reflects the size of the test set. The best operating point on this ROC curve is found at a NFR of 0.33 (requiring two thirds of the detected frames to be classed as abnormal before returning an abnormal diagnosis) resulting in sensitivity and specificity of 100% for simulated data and around 80% (83% TPR, 22% FPR) for real data. Other choices of MD threshold resulted in sensitivity-specificity combinations in the range (68%-90%) to (85%-70%).

VI. Concluding Remarks

We have presented a complete framework that is able to infer the 3D motion of carpal bones from a single view fluoroscopic sequence. It uses a hybrid statistical model to estimate both the pose and bone shapes from the fluoroscopic sequences allowing the motion of carpal bones during radial-ulnar deviation to be estimated. The positions and orientations in the image plane are estimated with high accuracy, and with slightly less accuracy in the out-of-plane direction. More importantly, the relative positions of the carpal bones can be estimated accurately. This is useful for detection of dissociation conditions. As an example of clinical application for this

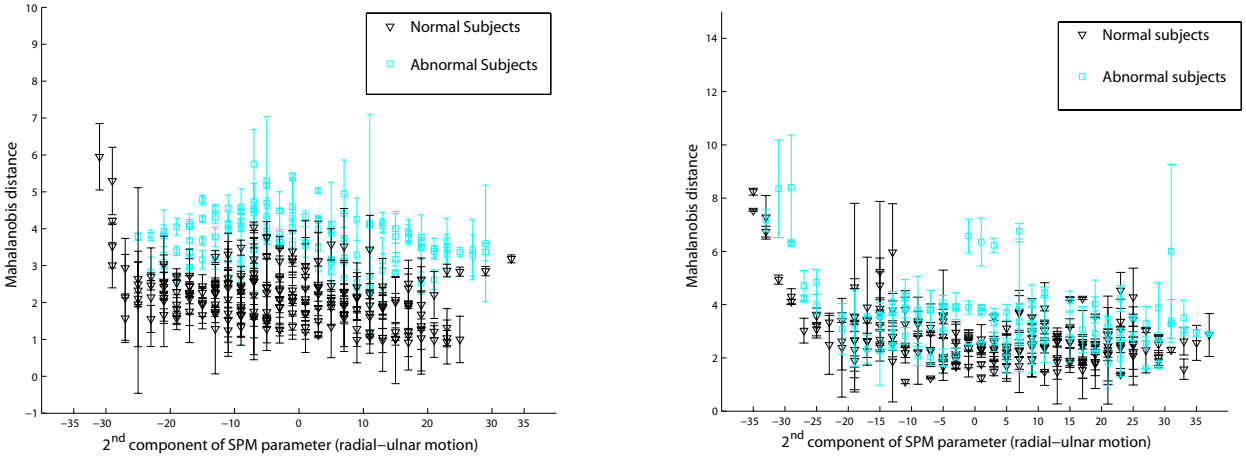

Fig. 8. (a) The Mahalanobis distances of 25 simulated sequences for Scaphoid-Lunate measurement. (b) The Mahalanobis distances of 15 real sequences for Scaphoid-Lunate measurement.

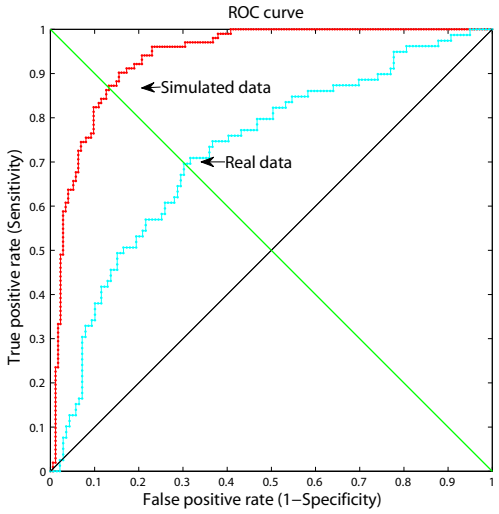

Fig. 9. ROC curve of the simulated data and real data for frame classification.

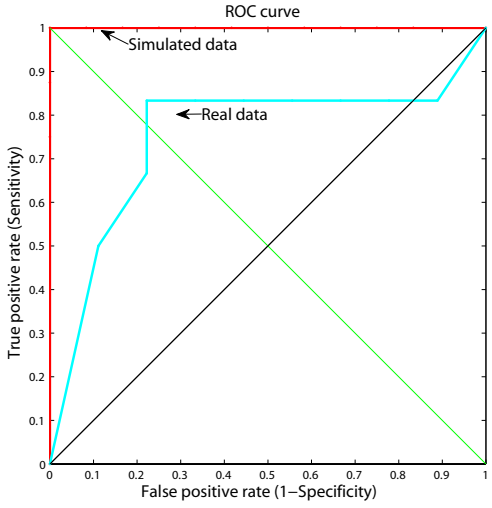

Fig. 10. ROC curve of the simulated data and real data for subject diagnosis.

type of analysis, we have used Scaphoid-Lunate dissociation, where the underlying pathology is a rupture of one or more ligaments, and the diagnosis rests on a judgement regarding the bone separation.

The proposed framework was tested on both simulated (25 subjects) and real (15 subjects) fluoroscopic sequences in the leave-one-out manner. The average absolute 3D point to point registration error is 2.45 $\pm$1.07 mm, whereas the errors along the in-plane directions, X^M and Z^M, average about 0.5 mm. There have been no comparable studies reporting cross-subject 2D-3D registration of multiple objects based on a single view. For comparison, [14] and [15] estimated the shape of the femur (a much larger structure) based on biplane X-ray images, reporting root mean square errors of 1.48 $\pm$ 0.41 mm and 1.4mm respectively. In our case, the relative 3D distances between bones can be estimated more accurately. The error of the measured 3D Scaphoid-Lunate distances are 0.75 $\pm$ 0.50 mm for simulated data and 0.93 $\pm$ 0.47 mm for real data. In addition, the reliability of the registration can be estimated by comparing the deviation of each bone from the SPM model poses with the deviations of other frames in the same sequence.

We also proposed, and conducted a preliminary evaluation of a method for constructing a 'standard' pathology measurement tool for automatically detecting Scaphoid-Lunate dissociation conditions, based on single-view fluoroscopic sequences. For the simulated data, it produced 100% sensitivity and specificity. For the real data, it achieved 83% sensitivity and 78% specificity. This tool could be a generic method for automatic, objective assessment of dissociation conditions. We have demonstrated its use with fluoroscopic video input. It appears that the limitation in accuracy arises largely from motion blurring effects in the video sequences. The method could equally well be applied using 2D radiographs at fixed positions. In a clinical setting, specified poses could be obtained using a fixation device.

The computational time for 1 frame was about 3 minutes

running in Matlab on a 3.6 GHz machine. For a typical sequence, this would result in three to five hours of computation, which would be acceptable for an off-line automatic analysis tool. If real-time feedback were required, faster computation would be necessary, which could be achieved by coding key parts in a compiled language, or use of GPU processing to parallelise the optimisation process.

In further work, we will extend the current statistical model with more training data (in progress), and improve the measurement model by including more healthy subjects. A larger training set may allow us a different compromise between constrained model fitting and local refinement. Here we have sought to avoid local minimal by restricting the SPM to only two modes of variation, relaxing the fit by local refinement. A larger dataset may result in a more specific model, making greater use of observed variability, reducing the need for the local refinement stage. However, if a range of abnormal conditions were to be included, the size of the training set might be prohibitive, requiring the retention of the local refinement. Our experience in this study indicates that it is a useful step in model fitting. On the basis of more data, we could further explore the relationship potentially associating the poses and shapes of bones. Nakamura et al. [24] have shown that carpal movement is affected by variation in the shape of the lunate. This raises the possibility that there may be more general relationships between bone shape and kinematics. It may be possible to build a more compact model by learning these relationships. We also intend to extend the framework to the (even) more challenging lateral views of flexion-extension motion, and further interpret the quantitative results for other wrist conditions. Acquiring a larger data set would also enable us comprehensively test the classification performance. Currently training and evaluation are conducted using the same data in a leave-one-out fashion.

ACKNOWLEDGMENT

The authors are grateful for the helpful suggestions of the anonymous referees, which have contributed to this paper.

REFERENCES

[1] J. G. Snel, H. W. Venema, T. M. Moojen, M. Ritt, C. A. Grimbergen and G. J. den Heeten, *Quantitative in vivo analysis of the kinematics of carpal bones from three-dimensional CT images using a deformable surface model and a three-dimensional matching technique*, Medical Physics, 27, pp. 2037-2047, 2000.

[2] S. E. Sonenblum, J. J. Crisco, L. Kang and E. Akelman, *In vivo motion of the scaphotrapezio-trapezoidal (STT) joint*, Journal of Biomechanics, 37, pp. 645-652, 2004.

[3] M. van de Giessen, G. J. Streekstra, S. D. Strackee, M. Maas, K. A. Grimbergen, L. J. van Vliet and F. M. Vos, *Constrained Registration of the Wrist Joint*, IEEE Transactions on Medical Imaging, 28(12), pp. 1861-1869, 2009.

[4] M. van de Giessen, M. Fournani, F. M. Vos, S. D. Strackee, M. Maas, L. J. van Vliet, K. A. Grimbergen and G. J. Streekstra, *A 4D statistical model of wrist bone motion patterns*, IEEE Transactions on Medical Imaging, 31(3), PP. 613-625, 2012.

[5] M. Foumani, S. D. Strackee, R. Jonges, L. Blankevoort, A. H. Zwinderman, B. Carelsen and G. J. Streekstra, *In-vivo three-dimensional carpal bone kinematics during flexion-extension and radio-ulnar deviation of the wrist: Dynamic motion versus step-wise static wrist positions*, Journal of Biomechanics, 42, pp. 2664-2671, 2009.

[6] C. Davatzikos, X. Tao and D. Shen, *Hierarchical active shape models, using the wavelet transform*, IEEE Transactions on Medical Imaging, vol 22(3), pp. 414-423, 2003.

[7] J. J. Cerrolaza, A. Villanueva and R. Cabeza, *Hierarchical statistical shape models of multiobject anatomical structures: application to brain MRI*, IEEE Transactions on Medical Imaging, vol 31(3), pp. 713-724, 2012.

[8] J. Boisvert, F. Cheriet, X. Pennec, H. Labelle and N. Ayache, *Geometric variability of the scoliotic spine using statistics on articulated shape models*, IEEE Transactions on Medical Imaging, vol 27(4), pp. 557-568, 2008.

[9] P. Marklj, D. Tomazevic, B. Likar and F. Pernus, *A review of 3D/2D registration methods for image-guided interventions*, Medical Image, Analysis, vol 16, pp. 642-661, 2010.

[10] G. P. Penney, P. G. Batchelor, D. L. G. Hill, D. J. Hawkes and J. Weese, *Validation of a two- to three-dimensional registration algorithm for aligning preoperative CT images and intraoperative fluoroscopy images*, Medical Physics, 28, pp. 1024-1032, 2001.

[11] D. B. Russakoff, T. Rohlfing, K. Mori, D. Rueckert, A. Ho, J. R. Adler Jr. and C. R. Maurer Jr., *Fast generation of digitally reconstructed radiographs using attenuation fields with application to 2D-3D image registration*, IEEE Transactions on Medical Imaging, 24(11), pp. 1441-1454, 2005.

[12] G. Zheng, *Statistically deformable 2D/3D registration for accurate determination of post-operative cup orientation from single standard X-ray radiograph*, MICCAI, pp. 820-827, 2009.

[13] T. Whitmarsh, L. Humbert, M. De Craene, L. M. Del Rio Barquero and A. F. Frangi, *Reconstructing the 3D shape and bone mineral density distribution of the proximal femur from dual-energy X-ray absorptiometry*, IEEE Transactions on Medical Imaging, 30(12), pp. 2101-2114, 2011.

[14] N. Baka, M. de Bruijne, T. van Walsum, B. L. Kaptein, J. E. Giphart, M. Schaap, W. J. Niessen and B. P.F. Lelieveldt, *Statistical shape model based femur kinematics from biplane fluoroscopy*, IEEE Transactions on Medical Imaging,31(8), pp. 1573-1583, 2012.

[15] G. Zheng, S. Gollmer, S. Schumann, X. Dong, T. Feilkas and M. A. G. Ballester, *A 2D/3D correspondence building method for reconstruction of a patient-specific 3D bone surface model using point distribution models and calibrated X-ray images*, Medical Image Analysis, vol 13, pp. 883-899, 2009.

[16] X. Chen, J. Graham, C. E. Hutchinson and L. Muir, *Inferring 3D kinematics of carpal bones from single view fluoroscopic sequences*, MICCAI 2011, volume 6892/2011, 680-687, 2011.

[17] X. Chen, J. Graham and C. E. Hutchinson, *Integrated framework for simultaneous segmentation and registration of carpal bones*, the 18^{th} ICIP, pp. 433-436, Belgium, 2011.

[18] J. J. Craig, *Introduction to Robotics: Mechanics and Control*, Addison-Wesley Publishing Company, Inc., 1989.

[19] R. H. Davies, C. Twining, T. F. Cootes and C. J. Taylor, *Building 3-D Statistical Shape Models by Direct Optimisation*, IEEE Transactions on Medical Imaging, 29(4), pp. 961-980, 2010.

[20] N. Amenta, *The Crust Algorithm for 3D Surface Reconstruction*, In: Proceeding of the fifteenth annual syposium on computational geometry, 1999.

[21] M. J. Black and G. Sapiro, *Edges as Outliers: Anisotropic Smoothing Using Local Image Statistics*, Scale-space theories in computer vision, 1682, pp. 259-270, 1999.

[22] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery *Numerical recipes in C++*,2nd Edition, Cambridge University Press, 2002.

[23] http://personalpages.manchester.ac.uk/staff/xin.chen/CarpalReg.htm

[24] K. Nakamura, M. Beppu, R. M. Patterson, C. A. Hanson, P. J. Hume and S. F. Viegas, *Motion analysis in two dimensions of radial-ulnar deviation of type I versus type II lunates*, Journal of Hand Surgery, vol 25(5), pp. 877-888, 2000.