3D Kinematics Estimation from Fluoroscopy Sequences for Wrist Pathology Diagnosis

Xin Chen[1]
xin.chen@manchester.ac.uk
Jim Graham[1]
jim.graham@manchester.ac.uk
Charles Hutchinson[2]
Lindsay Muir[3]

[1] Centre for Imaging Sciences
University of Manchester, UK
[2] Division of Health Sciences
University of Warwick, UK
[3] Department of Hand Surgery
Salford Royal NHS Foundation Trust, UK

Abstract

We present a novel tool for wrist pathology diagnosis by 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. We propose a method for constructing a 'standard' pathology measurement tool for automatically detecting Scaphoid-Lunate dissociation conditions. Evaluation on simulated fluoroscopic sequences produced 100% sensitivity and specificity. Evaluation on real fluoroscopic sequences achieved 83% sensitivity and 78% specificity.

1 Introduction

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. No tendons insert onto the carpal bones themselves, and their movements are therefore dictated by the movements of the surrounding bones. Knowledge of the 3D configurations of the bones in the wrist (carpal bones, radius and ulna) can lead to diagnosis of soft-tissue injury. However, clinically it is infeasible to apply 3D imaging (e.g. CT, MR) routinely in such cases. In particular, these are static images of a dynamic problem. The current method of diagnosing these conditions is by examining 2D video fluoroscopy sequences showing movement of the hand from full ulnar to full radial deviation and from full flexion to extension in two orthogonal views, in conjunction with plain-film radiographs at specific poses (stress views). From these images clinicians can infer the three-dimensional translations and rotations of the carpal bones that take place during wrist movement, and arrive at a differential diagnosis on the basis of variations from normal bone kinematics. The interpretation is difficult and the accuracy of diagnosis depends wholly on the experience of the practitioner. If the 3D bone positions could be determined automatically from 2D clinical radiographs, the diagnosis of associated soft-tissue injury, could be significantly improved.

A number of studies have represented the carpal kinematics using CT or MR data [2, 6]. More recently, van de Giessen et al. [7] introduced a 4D statistical model that locally describes the relative positions of the carpal bones in pre-defined poses, with the aim of detecting abnormal bone spaces. The work we present here is different from this, in that we intend to detect abnormalities using 2D fluoroscopic sequences rather than 3D volume data sets. In our previous work [3], we have described the method of estimating the 3D kinematics from AP view 2D fluoroscopic sequences. In this previous framework, a 3D statistical pose model (SPM) and statistical shape model (SSM), which were built from 25 subjects each at 3 radial-ulnar poses, are used to match with the 2D images by iteratively finding the optimum pose. In this paper, we extended the statistical training data to cover flexion-extension movement and build a unified shape model for all bones. More importantly, we demonstrate the capability of using the estimated 3D kinematic poses for wrist pathology diagnosis.

2 3D Kinematics Estimation from Fluoroscopy Sequences

Using the method described in [4], we generated a SPM based on transformation parameters of each bone with respect to a common reference coordinate system, as well as a SSM point distribution model. We extend the model building in [4], by using all the five poses (neutral pose and four extreme poses in flexion-extension and radial-ulnar deviations) from 25 subjects to cover the full range of wrist motion. Additionally, instead of building SSM for each individual bone, we build a single SSM that include all bones by representing the shape points of all bones in a single column vector in a consistent order. This maintains the nature of the relationships between adjacent bone shapes and reduces the number of shape parameters. In our experiments, only the first two significant components of the SPM are used, which keeps 90% of variation. The first component reflects the flexion-extension motion and the second component represents the radial-ulnar motion. The remaining deviation of an individual from the pose model is compensated by a local pose refinement for each individual bone. Based on the SPM and the SSM, a hybrid statistical mesh model can be built by using the Crust mesh construction algorithm [1]. Figure 1 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.

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. 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. The fluoroscopic image is iteratively matched with a simulated projection generated from an updated pose of the mesh model. The cost function for optimising the global parameters for rigid alignment, the SPM and SSM is the same as that described in [3], denoted E_1, and consists of the normalised correlation between the projected and fluoroscopic gradient images. The cost function is modified in the local refinement procedure (Equation (1)), by adding a term that keeps the estimated pose close to the pose model, preserving the topology of the carpal bones when the intensity term E_1 is weak. The local refinement accommodates deviations from the linear motions imposed by the linear SPM model.

$$E = E_1 + \omega \exp\left(-\frac{\frac{1}{p}\sum_{i=1}^{p} ||x_i^g - T^l(x_i^g)||^2}{2\sigma^2}\right) \tag{1}$$

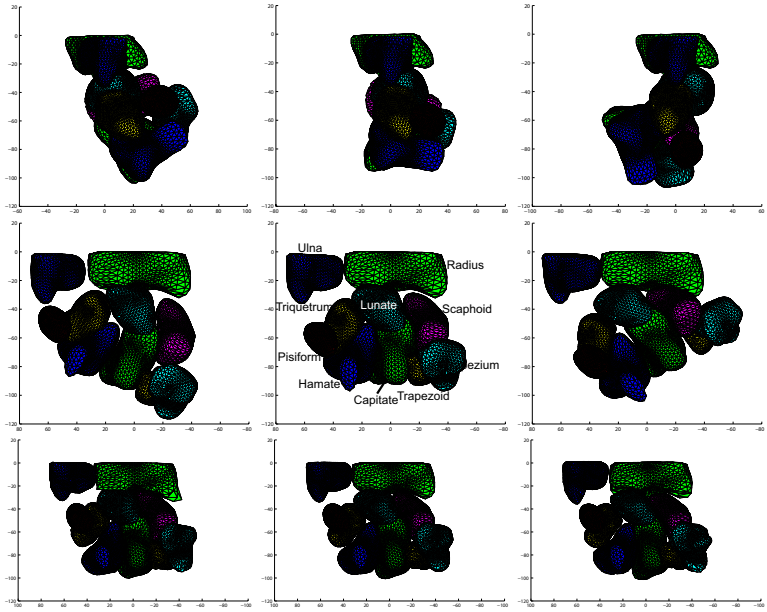


Figure 1: Top row: The poses of the first component of the pose model (lateral view) that mainly describes the flexion-extension movement. Middle row: The poses of the second component of the pose model (AP view) that mainly represents the radial-ulnar movement. Bottom row: the first component of the shape model. (Major shape variations occur in the Ulna, Radius and Lunate.) In each case the mean $\pm$ 2 s.d. are shown.

In equation (1), x_i^g represents the i^{th} 3D mesh point 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. σ is the standard deviation of the Gaussian distribution. In our evaluation tests, $\omega = -0.2$ and $\sigma = 10$ were experimentally determined and used.

3 Measurement Model for Pathology Detection

In [3], we have reported the relative positions of the carpal bones with respect to each other can be estimated with an accuracy of about 1 mm. Here, we present a measurement tool based on the estimated 3D wrist poses for pathology detection. One condition that may be assessed using the measurement of relative bone distances is dissociation, where the 3D distance between the bones is larger than normal. As an example of this, we investigate Scapho-lunate dissociation, which is one of the most common of these conditions.

One 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 pose determined by the kinematic model (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 (2).

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

In equation (2), x_i^g, p and T^l have the same meanings as in equation (1). Then the value r is subtracted from the mean deviation $\bar{r}$ 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).

The 3D CT volumes of 17 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. 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. Each bone is represented by the same number of surface points (1002), determined when the shape model was constructed using the minimum description length method [5]. Correspondences are determined using the index of each point, giving a consistent set of correspondences. Here we evenly sampled the points and used a reduced number of surface points (N=100, rather than 1002 used in building the model) for improved computational efficiency. Equation (3) and (4) show that we calculate the Mahalanobis distances (MD) using the means and covariances of individual pairs of corresponding points. Letting $l_{\phi,j}^k$ and $s_{\phi,j}^k$ represent the j^{th} surface point on the k^{th} sample volume at the current pose ϕ on the lunate and scaphoid respectively, the relative distance between the lunate and scaphoid at point j is

$$d_{\phi,j}^k = l_{\phi,j}^k - s_{\phi,j}^k \tag{3}$$

$d_{\phi,j}^k$ 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 (4).

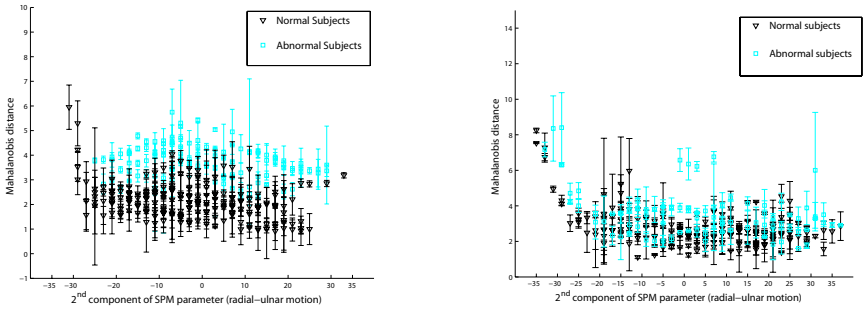
$$m_\phi = \frac{1}{N}\sum_{j=1}^{N} \sqrt{(d_{\phi,j}^{new} - m_{\phi,j})^T C_{\phi,j}^{-1} (d_{\phi,j}^{new} - m_{\phi,j})} \tag{4}$$

To assess a new wrist, the 2D radial-ulnar fluoroscopic sequence can be registered with the statistical model using the method described in section 2, and the wrist poses determined by the SPM component. The Mahalanobis Distance (MD) can then be calculated (Equation (4)) at each pose ϕ 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 2. In this figure the triangles represent healthy subjects and the squares represent abnormal subjects. In conducting this evaluation, a leave-one-out method was used in assessing the normal subjects, while the complete model was used for assessing the abnormals, as they did not contribute to the model. Since all 25 subjects contributed to the SSM, leave-one-out evaluation was used in all cases. The lengths of the bars through the data points represent the reliability of each registration, as calculated in Equation (2).

4 Results and Discussion

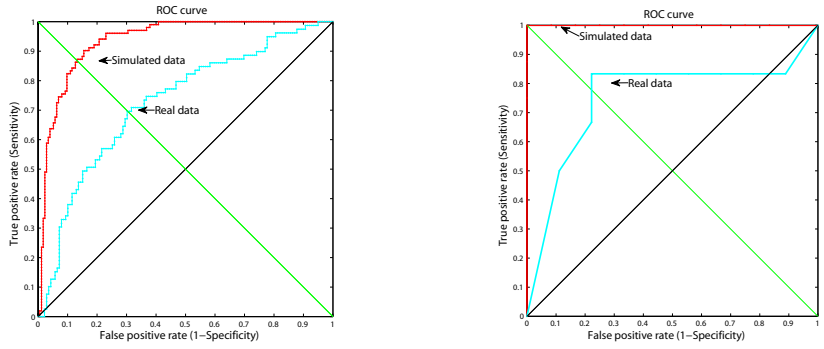
As shown in figure 2, for the simulated data, most of the abnormal subjects (squares) have larger MDs than the normal subjects (triangles). The distinction between the two groups is less pronounced for the real fluoroscopic sequences. Additionally, the registration is less reliable compared with the 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 3(a). The ROC for both the simulated data and real data are presented, using only the successful registrations (Eqn. (2)). 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.



(a) (b)

Figure 2: (a) The Mahalanobis distances of 25 simulated sequences for Scaphoid-Lunate measurement. (b) The Mahalanobis distances of 15 real sequences for Scaphoid-Lunate measurement.



(a) (b)

Figure 3: (a) ROC curve of the simulated data and real data for frame classification. (b) ROC curve of the simulated data and real data for subject diagnosis.

Figure 3(a) represents the accuracy of classification of individual frames. 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 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 3(b) 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%).

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 major contribution of this paper is that we conducted a preliminary evaluation of a method for constructing a pathology measurement tool for automatically detecting Scaphoid-Lunate dissociation conditions. 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. We would investigate the use of the model for diagnosis of other wrist diseases in our future work.

References

[1] N. Amenta. The crust algorithm for 3d surface reconstruction. In *15th Annu. Symp. Computat. Geometry*, pages 423–424, 1999.

[2] X. Chen, J. Graham, and C. Hutchinson. Integrated framework for simultaneous segmentation and registration of carpal bones. In *Proc. 18th ICIP*, pages 433–436, 2011.

[3] X. Chen, J. Graham, C. E. Hutchinson, and L. Muir. Inferring 3d kinematics of carpal bones from single view fluoroscopic sequences. In *MICCAI*, pages 680–687, 2011.

[4] X. Chen, J. Graham, C. E. Hutchinson, and L. Muir. 3d motion estimation of carpal bones from single view fluoroscopic sequences. In *MIUA*, pages 153–158, 2012.

[5] R. Davies, C. Twining, T. Cootes, and C. Taylor. Building 3-d statistical shape models by direct optimisation. *IEEE TMI*, 29(4):961–980, 2010.

[6] M. van de Giessen, G. Streekstra, S. Strackee, M. Maas, K. Grimbergen, L. van Vliet, and F. Vos. Constrained registration of the wrist joint. *IEEE TMI*, 28(12):1861–1869, 2009.

[7] M. van de Giessen, M. Fournani, F. Vos, S. Strackee, M. Maas, L. van Vliet, K. Grimbergen, and G. Streekstra. A 4d statistical model of wrist bone motion patterns. *IEEE TMI*, 31(3):613–625, 2012.