3D motion estimation in carpal bones from single-view fluoroscopic sequence

3D Motion Estimation of Carpal Bones from Single View Fluoroscopic Sequences

Xin Chen[1]
xin.chen@manchester.ac.uk
Jim Graham[1]
jim.graham@manchester.ac.uk
Charles Hutchinson[2]
c.e.hutchinson@warwick.ac.uk
Lindsay Muir[3]
lindsay.muir@srft.nhs.uk

[1] Imaging Sciences Research Group
University of Manchester
Manchester, UK

[2] Clinical Sciences Research Institute
University of Warwick
Coventry, UK

[3] Consultant Orthopaedic Surgeon,
Salford Royal Hospital
Salford, UK

Abstract

We present a novel framework for inferring 3D carpal bone kinematics and bone shapes from a single view fluoroscopic sequence. A hybrid statistical model representing both the kinematics 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 kinematics 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. We have evaluated the proposed method on both simulated data and real fluoroscopic sequences. It was found that the relative positions between carpal bones can be accurately estimated, which is potentially useful for detection of conditions such as scapholunate dissociation.

1 Introduction

The diagnosis of wrist pain is frequently achieved by inspection of 2D video fluoroscopic sequences showing movement of the hand in radial-ulnar or flexion-extension motion. This qualitative interpretation requires the inference of the 3D movement of the carpal bones from the 2D sequences, and requires considerable experience from the practitioner. We seek to achieve a quantitative analysis by computer interpretation. During wrist movement, the eight carpal bones follow a complex, multi-dimensional trajectory, making interpretation of radiographs difficult. For this study we have trained a hybrid statistical model (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 fluoroscopy sequences through 3D-2D image registration techniques. A number of studies have sought to represent the carpal kinematics using CT or MR data. Van deGiessen et al. [6] presented a 3D rigid registration method based on segmented meshes, which aims to build SM of carpal bones. More recently, they introduced a 4D statistical motion model that locally describes the movement patterns of the carpal bones [7]. 3D-2D registration has been the subject of many studies (e.g. [5]), mainly in the field of registration of pre-operative MR or CT images to intra-operative 2D

images. Our work differs from the above in that we seek to achieve registration of a 2D image sequence to a 3D model (not derived from the same individual) to derive the kinematics of an individual wrist.

The main contributions of this paper, distinguishing it from the aforementioned studies, are: (1) A hybrid SM is developed representing both the complex kinematics 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. We are not aware of any study which attempts to make a 2D to 3D inference in a system of this level of complexity.

The system consists of a training phase and a 3D-2D image registration phase. We currently have CT data from 10 subjects, each at five poses (neutral pose and two extreme poses in flexion-extension and radial-ulnar deviation). In the training phase, only the data from the neutral pose and two extreme poses in the radial-ulnar movement were used, as the radial-ulnar movement is the current concern of this paper. The segmentation of each bone and rigid registration parameters that align bones at different poses within and across the subjects were obtained using an iterative segmentation and registration algorithm [2]. A hybrid statistical model, representing both the kinematics and shape variation, was built from the results of the segmentation-registration framework. The kinematic model was built based on the transformation parameters, while the segmentation result was used to build the statistical shape model for each individual bone. Three sets of parameters need to be estimated during image registration in order to interpret the true 3D motion of the carpal bones: (1) Rigid transformation parameters of the wrist and a global scale factor, denoted by θ=$\{tx, ty, tz, rx, ry, rz, s\}$. tx and tz are the in-plane translations in AP view, and ty is the out-of-plane translation. $r = [rx, ry, rz]^T$ denotes the bone orientations, represented by Rodrigues parameter [3]. 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 to the origin in the radius, and the global size of the bones. (2) Kinematic model parameter b^m that produces valid poses of the carpal bones during movement. (3) Shape model parameters b_i^q and scale factor s_i for each bone (i). In the 3D-2D image registration phase, the three sets of parameters were estimated in sequence from each frame of the fluoroscopy sequences. Detailed descriptions are given in the following sections.

2 Training of Kinematic Model and Shape Model

We use the six rigid transformation parameters for each bone to train the kinematic model. The common coordinate system for all poses has an origin at a specified point in the radius for one subject. The pose of one subject is described by (tx_1, ty_1, tz_1, rx_1, ry_1, rz_1, ..., tx_{10}, ty_{10}, tz_{10}, rx_{10}, ry_{10}, rz_{10})^t. (8 carpal bones, 1 radius and 1 ulna). The orientation parameter allows for a continuous description of the wrist movement. Then the kinematic model can be parameterised as,

$$M = \mu^m + \phi^m b^m \tag{1}$$

where the mean pose μ^m (m is a notation indicating the model parameters) and the principal subspace matrix ϕ^m are computed from 3 (poses)$\times$ 10 (subjects) training samples using PCA. The vector b^m represents the kinematic parameters that describe the pose of M along each principal direction. In our experiments, the first 8 significant components are used, which keeps 98% of variation.

The statistical shape model of each bone is a point distribution model, built using the segmented volume of the same training subjects. The 3D structure of each bone is described by a set of approximately 1000 points on the segmented surface. Correspondence between these points across subjects was established by the minimum description length algorithm [4]. The deformable shape model is then described as,

$$Q_i = \mu_i^q + \phi_i^q b_i^q \tag{2}$$

where μ_i^q and ϕ_i^q (q is a notation indicating the shape parameters) are the mean shape and the principal subspace matrix for the i^{th} bone. b_i^q is the shape model parameter to be estimated. In order to keep the complexity within limits, only the first 3 significant components are used which keeps 84% of variation.

Based on the point distribution model of each bone and the kinematic model, a hybrid statistical mesh model can be built by using the Crust mesh construction algorithm [1]. Figure 1 shows the poses of the first component of the kinematic model (represented by the mean shapes of each bone) and the shapes of the first component of the scaphoid.

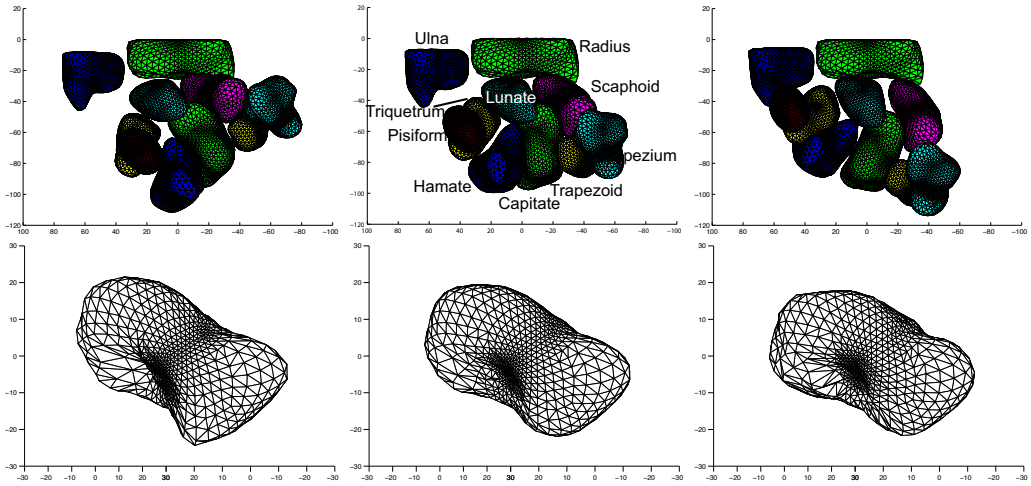

Figure 1: Top row: The poses of the first component of kinematic model. Bottom row: the first component of the shape model of the scaphoid. In each case the mean +/-1.5s.d. are shown.

3 3D-2D Image Registration

The statistical mesh model from the training data is then used to match with the fluoroscopic sequence to infer the 3D motion and bone shapes. Figure 2(a) summarises the registration process. Detailed descriptions are given in the following subsections.

3.1 Cost function

Preprocessing of the fluoroscopic image consists of normalising the local intensity to zero mean and unit standard deviation, followed by anisotropic diffusion. Figure 2(b) shows an example of the gradient magnitude map of the fluoroscopic image after enhancement. To optimise the pose parameters we iteratively generate projections from the mesh model with updated model parameters. Taking the mesh model as a binary volume, the projected intensity is negatively proportional to the sum of binary values along the ray from the source to each pixel in the image plane. Figure 2(b) shows an example.

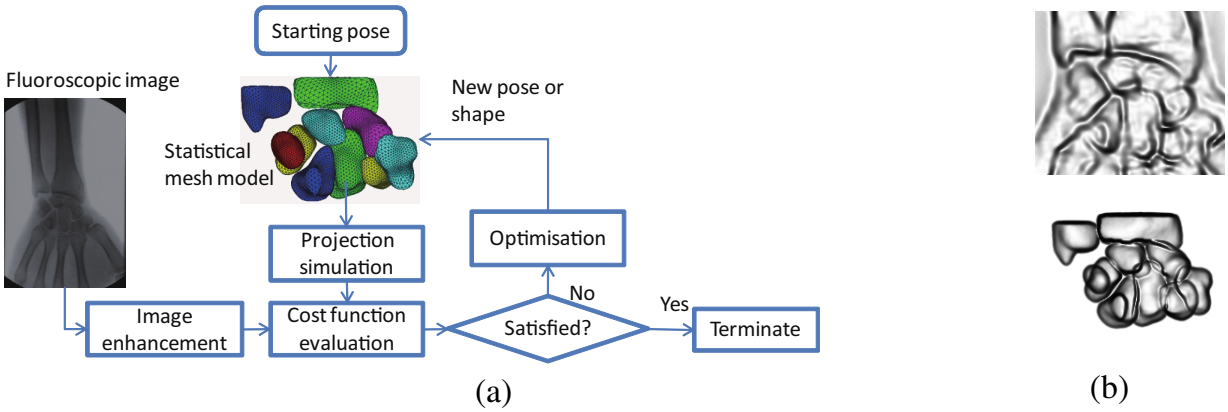

Figure 2: (a) Overview of the 3D-2D image registration process. (b) The gradient magnitude map of the fluoroscopic image after enhancement (cropped to show the region of interest) (top) and the simulated image from mesh model (bottom).

To evaluate the similarity between the fluoroscopic image and the simulated image, we use the cost function shown in Eqn. (3), based on the gradient along horizontal and vertical directions as well as the gradient magnitude of the two images. The adjacent frames of the current fluoroscopic image were also taken into account in the cost function to make the estimated poses smooth across frames.

Taking C(A,B) as the Normalised Correlation Coefficient between two images A and B, we can write the cost function as:

$$E = C(Om_{k-1}, Om_k) + \sum_{p=k-1,k,k+1} w_p(C(Im_p, Dm_k) + C(Ix_p, Dx_k) + C(Iy_p, Dy_k)) \quad (3)$$

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 second term calculates a cross-correlation between sets of three adjacent frames with weights w_{k-1}, w_k and w_{k+1}= 0.2, 0.6, 0.2 respectively, making the estimated pose smooth across frames. 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.

3.2 Optimisation

The optimisation method used is the best neighbour search combined with parabola fitting. The multi-dimensional search space (θ, b^m and b^q) is explored iteratively by individual 1D line search. The cost function is evaluated at the current position, positive and negative neighbour positions (defined by a search range), then an optimum is found by fitting a parabola to the 3 evaluated positions, and iteratively refined by reducing the search range until convergence.

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 out-of-plane translation, could be compensated by scaling of the bone. We estimate parameters (except ty) in the following sequence: tx, tz, rx, ry, rz, b^m, s, s_i and b_i^q. After convergence, the

estimated pose of the current frame is used as the starting pose for the next frame. The shape model parameters b_i^q are only estimated once in the first frame, as these are not improved significantly when we include more frames and the fitting is made significantly more complex and time consuming. The whole process was performed in a 3-level multi-scale framework at each frame to enhance the robustness of the registration.

4 Evaluation

As there is no ground truth relating the fluoroscopic sequences to the CT data, we evaluated our framework based on a number of simulated fluoroscopic sequences generated from the 3D CT data. All CT volumes have been resampled to an isocubic volume with voxel dimension of 0.5 mm. We linearly interpolated a number of poses between the neutral pose and two extreme poses of radial-ulnar deviation in a full movement cycle containing 50 poses for each of 10 subjects. The ray-casting method was then used to generate a simulated fluoroscopic sequence from those interpolated 3D data. We evaluated the proposed framework in the leave-one-out manner. The 3D pose of the simulated test subject was then calculated as described in section 3, and registration error measured by the 3D Euclidian distance of each corresponding point of the mesh between the target pose and the estimated pose is presented in Table 1. The error of the registration is mainly caused by the ill posed problem (confusion between the scale and translation along the out-of-plane direction), whereas the errors along the in-plane directions are very small with average error of about 1 mm and maximum error within 2mm.

It is important to mention that the relative positions of the carpal bones with respect to each other can be estimated much more accurately than the absolute positions of the individual bones. The registration error of the 3D distance between the centroid of Triquetrum and the centroid of Lunate (dTL), and the distance between the centroid of Lunate and the centroid of Scaphoid (dLS) were also measured. The errors are 0.59±0.37 mm and 0.91±0.50 mm for dTL and dLS respectively, compared to a bone size of about 15-20 mm. One of the conditions that may be assessed using this method is dissociation, where the distance between the bones is larger than normal. Scapholunate dissociation is one of the most common of these. We normalise the dLS by dividing it by the estimated global scale factor s and an average of the local scale factors for lunate and scaphoid. From the tested 10 subjects, we successfully identified the subjects suffering from scapholunate dissociation (dLS=19.39±0.77 mm) from the normal subjects (dLS=17.25±0.42 mm). Making this type of measurement without a 3D statistical model would be impossible.

Table 1: The average error in mm, measured in 3D, between the target and estimated correspondence points of each carpal bone of 10 subjects: Triquetrum(Tri), Lunate(Lun), Scaphoid(Sca), Pisiform(Pis), Hamate(Ham), Capitate (Cap), Trapezoid (Trd) Trapezium (Trm). The measurement errors of dTL and dLS.

	eTri	eLun	eSca	ePis	eHam	eCap	eTrd	eTrm	Total	eTL	eLS
Err3D	2.7±1.3	2.6±1.3	3.3±1.8	3.4±1.9	3.3±1.9	3.3±2.0	3.3±2.3	3.8±2.2	3.2 ±1.9	0.59±0.37	0.91±0.50
ErrX	0.8±0.7	1.0±0.8	1.1±0.9	1.2±1.0	0.9±0.7	1.1±0.8	0.9±0.7	1.1±0.9	1.1 ± 0.9	/	/
ErrY	1.9±1.4	1.5±1.3	2.5±1.9	2.4±2.0	2.7±2.1	2.7±2.2	2.8±2.5	3.1±2.3	2.3±2.0	/	/
ErrZ	1.3±1.0	1.3±1.0	1.1±0.9	1.3±1.1	1.3±0.7	0.9±0.7	0.8±0.6	1.1±1.0	1.2±1.0	/	/

We also tested our framework on real fluoroscopic sequences. Although the matching error cannot be quantified, the registration results show good visual correspondence and have been confirmed by a clinician. A sample frame of the matching result and the corresponding 3D pose are shown in Fig. 3 in which the projected contours from the 3D mesh model are superimposed on the preprocessed fluoroscopy image. The estimated 3D mesh model in the

palmar and dorsal views are shown in middle and right respectively.

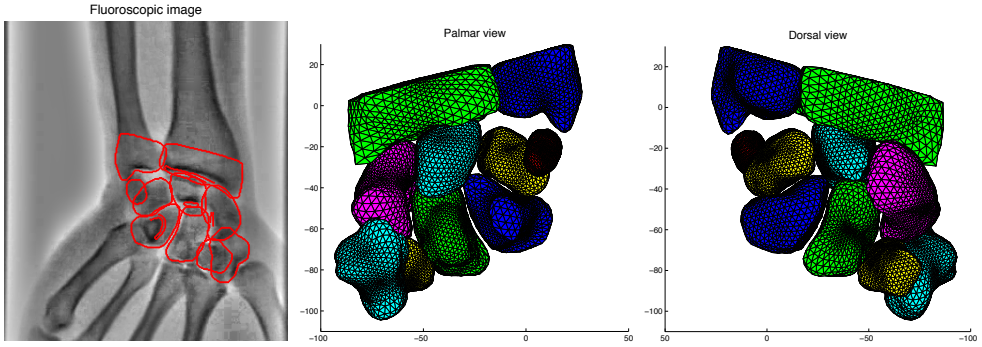

Figure 3: Registration result of one frame from a real fluoroscopic sequence.

5 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 kinematics and bone shapes from the fluoroscopic sequences allowing the motion of carpal bones during radial-ulnar deviation to be estimated. Particularly, the relative positions between carpal bones can be estimated accurately. This is potentially useful for detection of dissociation conditions, such as scapholunate dissociation, where the underlying pathology is a rupture of one or more ligaments, and the diagnosis rests on a judgement regarding the joint separation.

In further work we will extend the current statistical model with more training data (in progress) and test the framework for the flexion-extension movement.

References

[1] N. Amenta. The crust algorithm for 3D surface reconstruction. In *Proc. syposium on computational geometry*, 1999.

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

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

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

[5] 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*, 28:1024–1032, 2001.

[6] M. van deGiessen, 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 T-MI*, 28(12): 1861–1869, 2009.

[7] M. van deGiessen, M. Fournani, F. M. Vos, S. D. Strackee, M. Maas, L. J. van Vliet, K. Grimbergen, and G. J. Streekstra. A 4D statistical model of wrist bone motion patterns. *IEEE T-MI, online preprint, october*, 2011.