Correspondence free 3D statistical shape model fitting to sparse X-ray projections

N. Baka[a,b], W.J. Niessen[b,d], B.L.Kaptein[c], T. van Walsum[b], L. Ferrarini[a], J.H.C. Reiber[a], B.P.F. Lelieveldt[a,d]

[a] Division of Image Processing, Department of Radiology,
Leiden University Medical Center (LUMC), Leiden, the Netherlands
[b] Biomedical Imaging Group Rotterdam, Erasmus Medical Center, Rotterdam, the Netherlands
[c] Department of Orthopaedics, LUMC, Leiden, the Netherlands
[d] Delft University of Technology, Delft, the Netherlands

ABSTRACT

In this paper we address the problem of 3D shape reconstruction from sparse X-ray projections. We present a correspondence free method to fit a statistical shape model to two X-ray projections, and illustrate its performance in 3D shape reconstruction of the femur. The method alternates between 2D segmentation and 3D shaoe reconstruction, where 2D segmentation is guided by dynamic programming along the model projection on the X-ray plane. 3D reconstruction is based on the iterative minimization of the 3D distance between a set of support points and the back-projected silhouette with respect to the pose and model parameters. We show robustness of the reconstruction on simulated X-ray projection data of the femur, varying the field of view; and in a pilot study on cadaveric femora.

Keywords: 2D/3D reconstruction, SSM, biplane X-ray, correspondence free, femur

1. INTRODUCTION

Three dimensional shape reconstruction from a sparse set of 2D projection images is important for a range of medical applications. Recently, image guided intervention systems were introduced in e.g. orthopedic surgery. These systems often require a patient specific 3D anatomical model, which should be related to the intra-operative situation. If a 3D CT dataset is not available, an alternative is to generate a 3D model from calibrated X-ray projections. In this case, since only small number of projections is available, integration of prior knowledge is required for successful 3D reconstruction, e.g. through the use of statistical shape models (SSM).[1–5]

Prior work on bony anatomy reconstruction based on sparse data can be arranged in two main groups according to SSM type: 1) statistical deformation models and 2) Point Distribution Models (PDM). Deformation models are based on statistics of the deformation fields converting one subject to the mean. They have been mainly applied for cross-sectional, or point based sparsity scenarios.[6,7] In case projection information is available, such as X-Rays, point distribution models are favored.[1–5] In the rest of this paper the names SSM and PDM are used as synonyms. The first 2D/3D non-rigid PDM fitting mechanism was introduced by Lavallee[8] casting the problem to 3D point cloud fitting to PDMs. The 3D position of a 2D edge point is defined as the point on the ray, which is closest to the model surface. A follow up work consisted of 3D point correspondence establishment for determining the point-to-surface distance.[4] Another approach employing 3D distances relies on prior 2D correspondence building between edges of the X-Ray, and the projected mean model. This 2D correspondence is mapped to 3D by selecting the closest ray point to the corresponding model point.[9]

Purely 2D similarity metrics were applied in[1] and.[2] A scheme was suggested to derive edge potentials from the X-ray images, and subsequently calculate the distance function by averaging the edge potentials along the projected contour lines.[1] A similar approach is presented for silhouettes , and applied to hip images.[2] A recently published method employs rigid registration inspired by the physical movement model of a rigid object described by local forces and moments. For non-rigid fitting the similarity measure calculated from the distance between the 2D segmentation and the projected model silhouette is minimized with respect to the model parameters.[5]

Generally, 2D based similarity metrics have the advantage of enabling correspondence free matching[1,2,5] , but have the disadvantage that projection distortion resulting from the 3D nature of the original shape, and perspective projection, cannot be corrected for. On the other hand approaches based on 3D distances eliminate the projective distortion, however both[4,9] require a computationally expensive point correspondence calculation.

Furthermore, in the methods mentioned above only the part of the bone is modeled that is present in the imaging Field of View (FOV), e.g. proximal or distal part of the femur. Often separate models are required to model subparts of the femur for different clinical applications.

In this work we focus on shape reconstruction from uni-planar stereo X-rays of femora for pre-operative planning of orthopedic surgery. Such a reconstruction from stereo X-rays could possibly replace the pre-operative CT scan. The technical contribution of this work is twofold:

1. We develop a *3D correspondence free* reconstruction method. This method recovers projective distortion, and eliminates the ambiguous and computation intensive correspondence building step. We furthermore reduce the overall number of cost-intensive silhouette calculations.
2. We demonstrate robustness of the method to partially missing data. In many situations only part of the shape is visible due to image size restrictions. We evaluate the effect of different field of views (FOV) using a femur model.

The paper is organized as follows. Section 2 gives a detailed explanation of the proposed algorithm. The data used for evaluation is presented in Section 3, followed by the description of the experimental setup in Section 4. Results and discussion is given in Section 5. Finally, Section 6 concludes the paper.

2. METHODS

Method Overview The presented method relies on the known geometry of X-ray sources and projection planes, i.e. we assume a calibrated imaging setup. Prior knowledge of the shape is described by a PDM.[10] The aim of the method is 2D segmentation as well as 3D reconstruction. Correspondingly, the algorithm consists of two main building blocks: the 2D segmentation given the 3D shape model, and the 3D reconstruction given the 2D segmentations on the X-Ray images. These two blocks are applied in an iterative manner. The description of the full cycle according to Figure 1 is as follows. After manual initialization of the 3D model (stage 1), the model is projected on the X-ray images (stage 2). By applying dynamic programming (stage 3), we derive an updated silhouette (stage 4) for each view separately. Using the pre-calculated measures (stage 5), 3D distances per view are efficiently calculated (stage 6). A 3D registration criterion based on these distances is minimized with regard to pose and shape (stage 7), resulting in a new shape estimate (stage 1). This loop is iterated either over a fixed number of iterations, or until convergence.

Model building In this work we use aPDM as prior information.[10] The statistical shape model is built from N 3-dimensional training shapes. Shapes are segmentations of training images. Subsequently, the GAMES algorithm[11] is applied to represent each shape with the same number n of corresponding landmarks. The landmarks are interconnected to form triangulated surface patches. In order to retain solely shape variations, the shapes are priorly aligned with Procrustes analysis. Each shape can be represented as a point x in the $n * d$ dimensional shape space. The set of training shapes $X = x^1, x^2, ..., x^N$ forms a point cloud in this space. Dimension reduction techniques like PCA are now applied resulting in a model of the form

$$x = \bar{x} + \Phi b + \epsilon \quad \longrightarrow \quad \tilde{x} = \bar{x} + \Phi b \, , \tag{1}$$

where $\bar{x}$ is the mean shape, Φ is the matrix of the main modes of variation, b is the representation of the shape in parameter space (where $b << n$), and ϵ is a small error due to the dimension reduction.

2D Segmentation The current 3D model is projected on each X-ray image. The resulting silhouette is sampled equi-distantly at positions P_j ($j \in [1, k]$). Scan lines normal to the silhouette are used to compute the gradient magnitude vectors f_j. In practice these vectors might lead outside the field of view (FOV). In such cases the vector is completed by filling the missing elements with a uniform value d, which is chosen such that dynamic programming will not favor positions outside the FOV. The feature matrix $F = [f_1 | \ldots | f_k]$ is constructed containing the feature vectors in the columns, representing the straightened shape and the directional derivative values in a given proximity to the contour. The correct edges in the X-Ray are characterized by high absolute derivative values, and a negative sign. The optimal contours are found via dynamic programming with side-step costs controlling the stiffness of the contour. Dynamic programming solves effectively the min cost path problem.[12] For a wider capture range, multiple scales can be adapted.

Finally, a likelihood η is attached to each pixel of the updated silhouette, derived from the corresponding gradient magnitude of the image.

3D reconstruction *Support point selection from a single view.* Support points are 3D model landmarks that are being used in the fitting process. These points are selected according to a 2D similarity criterion based on surface normal and distance. Points close to the back-projected contour with high similarity between detected 2D contour normal and projected 3D surface normal are selected as support points. Similarity is bounded by a maximum distance D_{max} and a maximum angle Θ_{max} as follows (see Fig. 2):
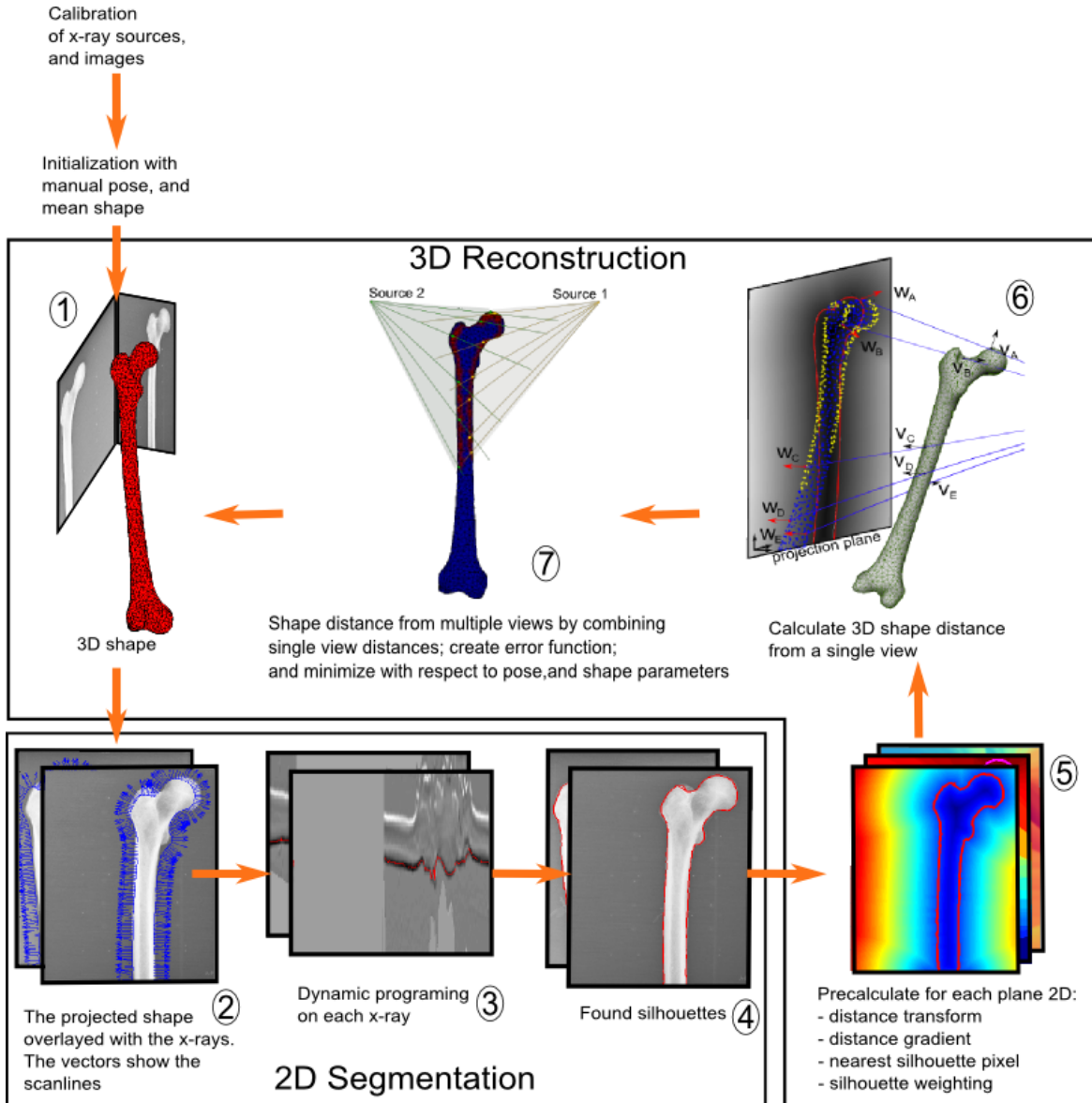
Figure 1. Iterative 2D segmentation and 3D reconstruction

1. The angle Θ_i between the normal vector $\vec{V}_i$ at each landmark point i and the estimated normal $\vec{W}_i$ at the projected landmark position is calculated. The estimated normal vectors are defined on the plane as the steepest gradient of the 2D distance transform of the silhouette at a particular pixel position ($\nabla \mathrm{DT}(x_i, y_i)$), and out of plane such that it is perpendicular to the ray connecting the source and the particular pixel in plane (r_i):

$$\vec{W}_i = \nabla \mathrm{DT}(x_i, y_i) + \frac{\nabla \mathrm{DT}(x_i, y_i) \cdot \vec{r}_i}{-\vec{e_z} \cdot \vec{r}_i} \vec{e_z} \qquad , \qquad (2)$$

$$\Theta_i = \mathrm{acos}\left(\frac{V_i W_i}{|V_i||W_i|}\right) \qquad . \qquad (3)$$

The landmark point i is selected, if $\Theta_i < \Theta_{max}$.
2. The 3D distance d_i of a selected landmark point to the closest silhouette ray is restricted by parameter D_{max}, increasing robustness with respect to outliers.
3. In the ideal case all landmark points fall inside the silhouette. Therefore, landmarks outside the silhouette are penalized by multiplying the distance with a factor $\rho > 1$. For inside points, $\rho = 1$.
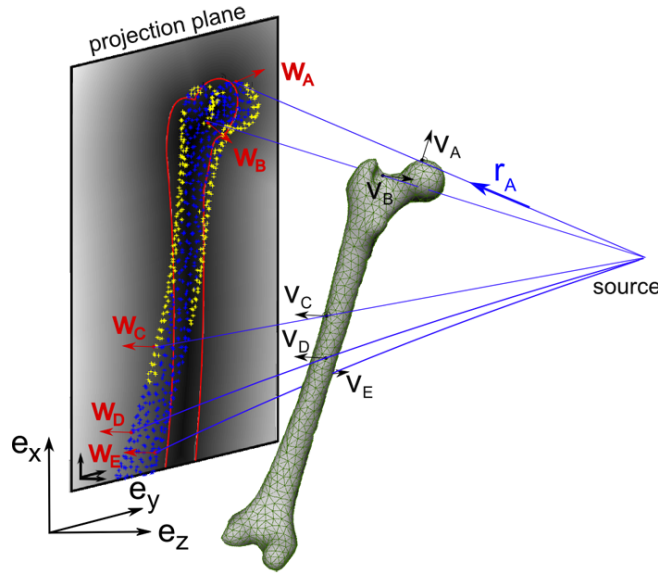
Figure 2. Support point selection: All landmark points are projected on the plane; subsequently landmarks are selected based on angle, and threshold criteria. Landmarks marked with yellow are selected as model support points. Landmark A, B and E have angles of approximately 90°, 90° and 180°, are thus not selected. Landmark D is not selected as it is too far from the silhouette.

4. Unreliable parts of the silhouette could mislead the match. For that reason all distances from silhouette positions $p \in Silhouette$ p are weighted with the underlying image gradient magnitude η at p.

Criterion Function and Minimization Scheme. The step from single projection to multiple projections is achieved by combining the support point sets of each of the projections. If a landmark i is a supported point in several projections, its distance D_i is the mean of the individual distances from all projections where it was selected.

The registration criterion function is defined as the mean square error over the set of support points. To favor statistically plausible deformations the Mahalanobis distance of the shape parameter vector to the mean shape is included in the final energy penalizing large deviations from the mean shape.

$$E_{PDM}(D_{max}, \Theta_{max}, \rho) = \frac{1}{n} \sum_{i=1}^{n} D_i^2 + \alpha (p - \mu)^T \Sigma^{-1} (p - \mu)$$

, where $i \in L_{support}$, α is a weighting factor, p is the shape parameter vector, μ and Σ are the mean and covariance matrix of the SSM.

Fitting proceeds as follows: After manual initialization, rigid parameters are optimized first. Further refinement includes optimizing shape parameters as well. Initially Θ_{max} and D_{max} are assigned large values, resulting in a large support point set and a coarse registration with a wider convergence radius. By gradually decreasing Θ_{max} and D_{max}, accuracy of the fit increases.Minimization of the error function is performed using conjugated gradients.

An example fit with a small field of view is shown in Figure 3. The figure shows three stages of the algorithm: initialization, affine fit, and final shape fit. The light blue area denotes the silhouette on the projection images, the model projection on the plane is drawn dark blue, and the overlap of these two is colored yellow. Selected points for a given angle and threshold are shown as red stars. The upper left image pair depicting the initial position illustrates well the need of a large angle and threshold for the landmark selection mechanism for the first iterations. A too small angle, and threshold may result in too few selected landmarks. In the example of Figure 3, only one landmark gets selected in the lateral projection at a maximal angle of 30°, and threshold of 20mm.

3. DATA

Model construction. The model was constructed using a semi-automatic segmentation method from CT Angiography scans of the lower extremities of 30 patients. Segmentations were obtained from 15 mirrored right, and 15 left femora. The population contained 23 male and 7 female subjects aging between 23 and 66 years. Patients were scanned
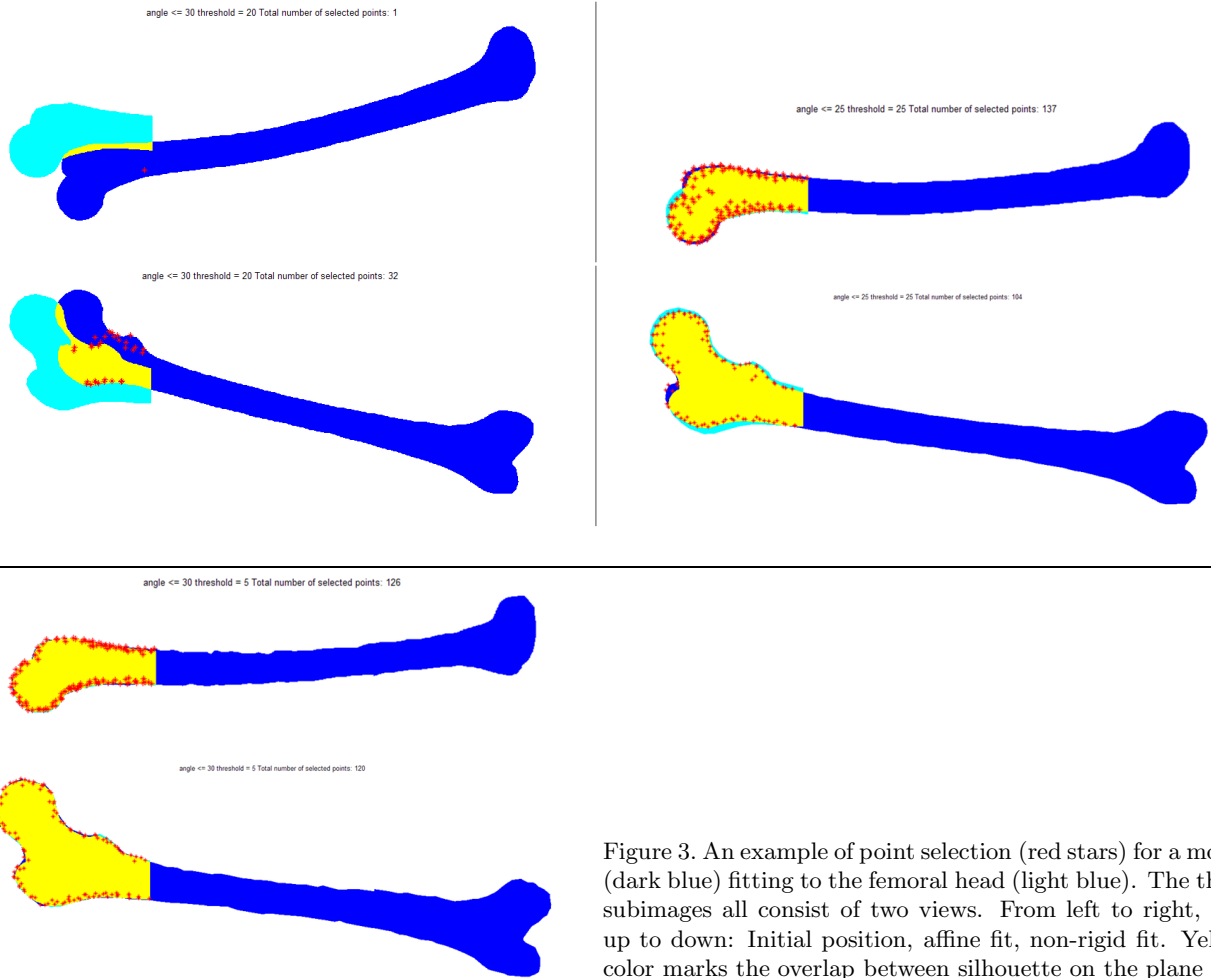
Figure 3. An example of point selection (red stars) for a model (dark blue) fitting to the femoral head (light blue). The three subimages all consist of two views. From left to right, and up to down: Initial position, affine fit, non-rigid fit. Yellow color marks the overlap between silhouette on the plane and projected model.

on different scanners with voxel sizes ranging from 0.625 x 0.625 x $0.8mm^3$ to 0.782 x 0.782 x $0.8mm^3$. Landmark points on the training shapes were calculated using the GAMES algorithm[11] with an average landmark distance of 3mm, yielding 931 landmarks for the entire femur. Retaining 95% variance resulted in 23 modes of variation.

Cadaver study. Uni-planar stereo X-rays of 2 cadaver femora with metal markers were acquired using a Roentgen Stereophotogrammetric Analysis (RSA) setup. The cadaver femora were placed on a calibration box, which enabled the computation of the exact acquisition geometry.[13] Two views of the femur were acquired simultaneously, with an angle of $35°$ containing approximately 50-60% of an average length femur. Each cadaver proximal femur was scanned 4 times, with different orientations (natural, endo, epi, back). For validation purposes the same cadaver femora (including the metallic markers) have been segmented from CT, to be used as ground truth.

4. EXPERIMENTAL SETUP

Experiments. We performed experiments to 1) determine the effect of the FOV on the reconstruction accuracy, and to 2) determine how well the 3D femur shape can be predicted in a clinically more realistic setup.
For the first experiment we performed tests on simulated biplane projections ($90°$) to evaluate performance under controlled conditions. A set of 100 plausible femora were randomly generated by sampling shape instances from the statistical shape model of the femur; parameters were varied following a normal shape-, and uniform pose-distribution (translation: $\pm 15mm$, rotation: $\pm 15°$, scaling: $0.9 - 1.1$). The initial pose was zero translation and rotation, unit scaling, and mean

shape. The reconstruction error of different regions was tested after fitting the whole femur model to the available information. Available information included 3 FOV sizes: the entire femur (100%), the femoral head and part of the shaft (proximal 60%), and the femoral head (proximal 30%).

Secondly, a cadaver study of fitting the 3D model to calibrated uni-planar stereo Xrays (subsampled to pixel size $0.48 \times 0.48 mm^2$) was repeated 30 times for each acquisition direction, with an initial pose randomly selected in a radius of $10mm$ translation, $10°$ rotation, and $0.9 - 1.1$ scaling around the true position. Optimization was performed in three rigid optimization loops, followed by three cycles of shape optimization with decreasing angle and distance thresholds.

Error metrics. The 3D reconstruction accuracy was evaluated using the average point-to-surface (P2S) error for the whole femur, the partial femur inside the FOV, and the femoral head. Matching convergence was defined as a mean P2S error smaller than $3mm$ inside the FOV.

The 2D segmentation accuracy was defined by the average silhouette error, which is the sum of all false negative and false positive pixels divided by the silhouette length of the desired segmentation. Two types of average silhouette errors were calculated. The *Average 2D* error is calculated between the projected final 3D shape, and the automatic segmentation. This indicates the ability of the 3D reconstruction to deduct a shape from the 2D contours. Secondly, the *Segmentation* error is given, which evaluates the 2D segmentation part of the algorithm, and is measured between the gold standard contours, and the automatic segmentation.

5. RESULTS

Quantitative results of the simulated experiments are given in Table 1. P2S errors of the 100% FOV illustrate the performance of the reconstruction when data support for the entire model is available. Errors mainly occur in occluded areas, as shown in Fig. 4. We see that the smaller the FOV, the higher overall P2S errors occur. Generally, when an entire subpart of the femur is missing, reconstruction becomes ambiguous. There is no information e.g. about the length of the bone, leading to translation uncertainties, as translation is defined by the position of the mid-point of the shape. Ambiguities arise also from rotation vs shear in the shape. It is thus not surprising, that the missing distal part of the femur was not reconstructed properly. However, the performance inside the FOV is comparable for all FOV sizes, showing the robustness of the presented algorithm to missing data.
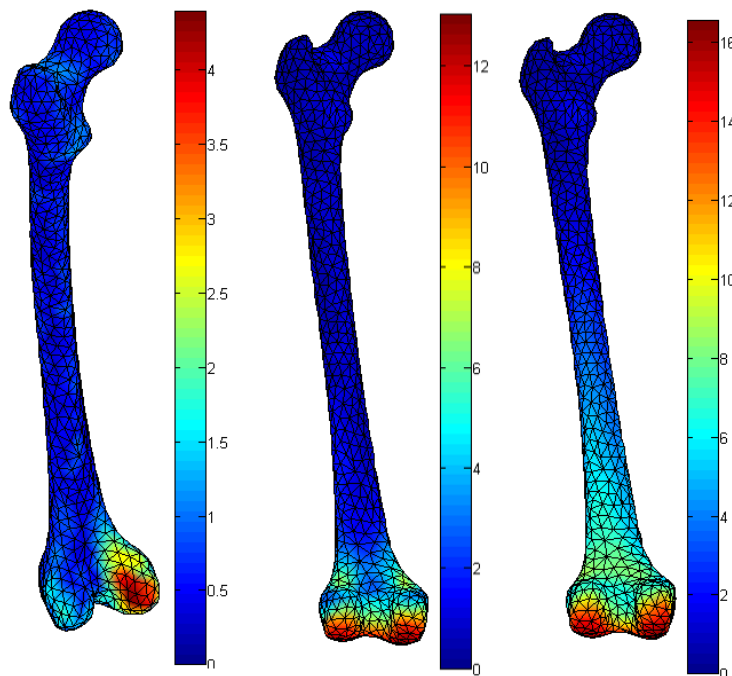

Figure 4. P2S error distributions along the femur, an average of all experiments. FOV sizes from left to right are 100%, 60%, and 30%.

Table 1. Results of the simulated experiments. First column indicates the FOV size used for fitting the entire femur model. All errors are given in mm.

Exp	Conv	P2S whole	P2S FOV	P2S Head	Average 2D	Segmentation
100%	100/100	0.75±0.28	0.75±0.28	0.6±0.11	0.63±0.22	-
60%	100/100	2.25±0.95	0.69±0.1	0.7±0.1	0.6±0.12	-
30%	98/100	4.13±1.45	0.76±0.21	0.82±0.42	0.63±0.25	-

Table 2. Results of the cadaver experiments. All errors are given in mm.The average 2D error shows the error between projected 3D reconstruction and the final segmentation derived from the dynamic programming. The column "Segmentation" denotes the error between final dynamic programming and projected ground truth femur.

Exp	Conv	P2S whole	P2S FOV	P2S Head	Average 2D	Segmentation
Cadaver1						
natural	30/30	7.48±1.41	1.61±0.25	2.15±1.63	1.02±0.14	1.19±0.08
epi	30/30	6.55±1.11	1.51±0.13	1.51±1.16	1.26±0.23	1±0.09
endo	30/30	6.56±2.0	2±0.29	1.8±1.24	1.1±0.14	1.17±0.09
back	30/30	7.47±1.49	2.2±0.97	1.55±1.16	1.42±0.58	0.64±0.05
Cadaver2						
natural	30/30	6.6±1.62	1.4±0.15	1.56±1.15	0.85±0.12	0.75±0.02
epi	30/30	6.35±2.18	1.22±0.12	1.44±1.05	0.86±0.09	0.97±0.18
endo	30/30	8.92±1.5	2.5±0.22	2.42±2.1	1.27±0.48	1.04±0.23
back	30/30	7.00±1.6	1.78±0.34	2.15±1.74	1.35±0.27	0.93±0.15

Figure 5 illustrates a fit of the model to a cadaver bone, showing the two uni-planar stereo X-Ray images overlayed with the initial position, affine registration result, and final fit result. Quantitative evaluation is given in Table 2. Compared to the simulation experiments, the accuracy of the cadaver studies decreased due to three factors: 1) the cadaver femora were from a different population than the model femora; 2) the angle between the projection planes was smaller than in the simulation experiments; 3) in the cadaver case, also segmentation errors contribute to the P2S reconstruction error (whereas in the simulation experiments, the perfect silhouette was employed for reconstruction).

Computing times ranged between 2 and 3 minutes in an unoptimized Matlab implementation. First results are promising, however further refinements are still necessary before clinical application is feasible.

The above evaluation was done using the model defined in Section 3, having the property of uniform landmark distribution along the surface. The landmark sampling can be optimized for this application, by making it curvature dependent. High curvature areas should be sampled at a denser landmark distribution than low curvature ones ensuring a balanced support point selection even at small maximal angles. A further way to improve the results is an overall denser sampling of the femur, and expansion of the training set.

6. CONCLUSIONS

Numerous publications tackle the problem of 3D bone reconstruction from biplane X-ray images.[1–5] However, these methods model only part of the bone visible in the imaging FOV (e.g. proximal, or distal). Often separate models are required to model subparts of the femur for different clinical applications. Furthermore, 2D based similarity metrics enabling correspondence free matching, do not correct for projection distortion.[1,2,5] Approaches based on 3D distances on the other hand eliminate the projective distortion, however both[4,9] require a computationally expensive point correspondence calculation.

Therefore, the contribution of our work is twofold: 1) we presented a new correspondence free 3D PDM fitting approach for surface reconstruction from a variable number of X-ray projections (> 1); 2) we demonstrate robustness of the algorithm with respect to reduced FOV. We evaluated our approach on simulated data using two projections with a femur model, and on cadaveric femora. The first experimental results are promising, further steps towards usage of the method in pre-operative planning of orthopaedic surgery will follow. Future work will focus on optimized model building for more accurate results. Also, the current 3D reconstruction method requires a connected silhouette segmentation on the projection images. In order to apply the same method on a wider range of shapes, we also consider adding inside edges to the fitting.
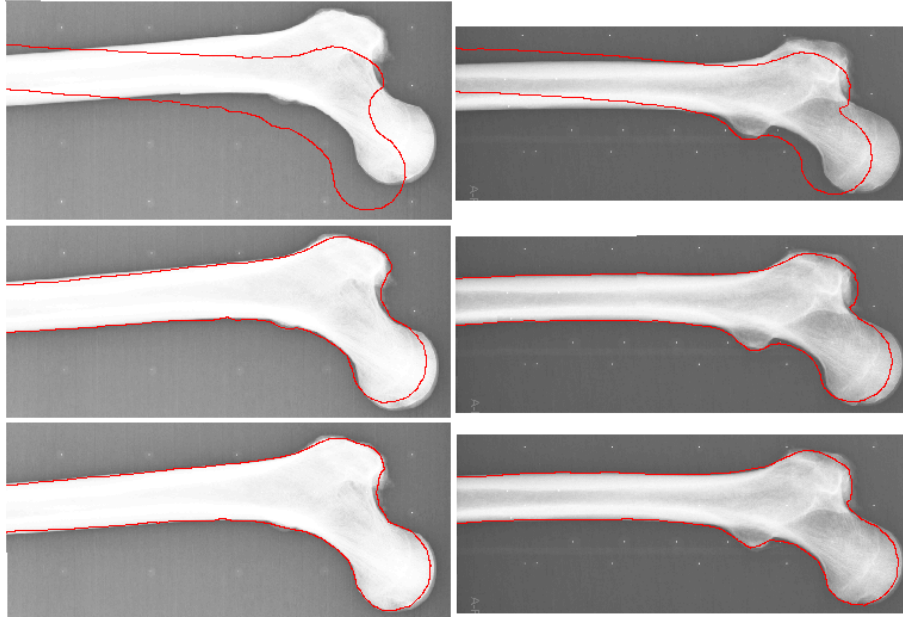
Figure 5. Cadaver2 neutral position. Top: initialization; Middle: rigid registration result; Bottom: Non-rigid matching result

REFERENCES

[1] S. Benameur, M. Mignotte, S. Parent, H. Labelle, W. Skalli, and J. D. Guise, "3D/2D registration and segmentation of scoliotic vertebrae using statistical models," *Computerized Medical Imaging and Graphics* **27**(5), pp. 321–337, 2003.

[2] H. Lamecker, T. H. Wenckebach, and H.-C. Hege, "Atlas-based 3d-shape reconstruction from x-ray images," in *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pp. 371–374, IEEE Computer Society, (Washington, DC, USA), 2006.

[3] G. Zheng and L.-P. Nolte, "Surface reconstruction of bone from x-ray images and point distribution model incorporating a novel method for 2d-3d correspondence," in *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2237–2244, IEEE Computer Society, (Washington, DC, USA), 2006.

[4] M. Fleute and S. Lavallée, "Nonrigid 3-d/2-d registration of images using statistical models," in *MICCAI '99: Proceedings of the Second International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 138–147, Springer-Verlag, (London, UK), 1999.

[5] R. Kurazume, K. Nakamura, T. Okada, Y. Sato, N. Sugano, T. Koyama, Y. Iwashita, and T. Hasegawa, "3d reconstruction of a femoral shape using a parametric model and two 2d fluoroscopic images," *Comput. Vis. Image Underst.* **113**(2), pp. 202–211, 2009.

[6] D. C. Barratt, C. S. Chan, P. J. Edwards, G. P. Penney, M. Slomczykowski, T. J. Carter, and D. J. Hawkes, "Instantiation and registration of statistical shape models of the femur and pelvis using 3d ultrasound imaging," *Medical Image Analysis* **12**(3), pp. 358 – 374, 2008.

[7] C. S. K. Chan, D. C. Barratt, P. J. Edwards, G. P. Penney, M. Slomczykowski, T. J. Carter, and D. J. Hawkes, "Cadaver validation of the use of ultrasound for 3d model instantiation of bony anatomy in image guided orthopaedic surgery.," in *MICCAI (2)*, C. Barillot, D. R. Haynor, and P. Hellier, eds., *Lecture Notes in Computer Science* **3217**, pp. 397–404, Springer, 2004.

[8] S. Lavallée, R. Szeliski, and L. Brunie, "Matching 3-d smooth surfaces with their 2-d projections using 3-d distance maps," in *Selected Papers from the Workshop on Geometric Reasoning for Perception and Action*, pp. 217–238, Springer-Verlag, (London, UK), 1993.

[9] 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* **13(6)**, pp. 883–899, 2008.

[10] T. F. Cootes, D. Cooper, C. J. Taylor, and E. Di Mauro, "Active shape models - their training and application," *Comput. Vis. Image Underst.* **61**, pp. 38–59, 1995.

[11] L. Ferrarini, H. Olofsen, W. Palm, M. van Buchem, J. Reiber, and F. Admiraal-Behloul, "Games: growing and adaptive meshes for fully automatic shape modeling and analysis," *Medical Image Analysis* **11**, pp. 302–14, 2007.

[12] R. Bellman and S. Dreyfus, *Applied Dynamic Programming*, Prinston Unicersity Press, 1962.

[13] E. Valstar, R. Gill, L. Ryd, G. FLivik, N. Börlin, and J. Kärrholm, "Guidelines for standadrization of radiostereo-ogrammetry (rsa) of implants," *Acta Orthop.* **76(4)**, pp. 563–72, 2005.