AUTOMATIC EXTRACTION OF ANATOMICAL LANDMARKS FROM MEDICAL IMAGE DATA: AN EVALUATION OF DIFFERENT METHODS

Heiko Seim[1*], *Dagmar Kainmueller*[1†], *Markus Heller*[2], *Stefan Zachow*[1], *Hans-Christian Hege*[1]

[1]Medical Planning Group, Zuse Institute Berlin, Germany
[2]Julius Wolff Institut and Center for Musculoskeletal Surgery, Charité, Berlin, Germany

ABSTRACT

This work presents three different methods for automatic detection of anatomical landmarks in CT data, namely for the left and right anterior superior iliac spines and the pubic symphysis. The methods exhibit different degrees of generality in terms of portability to other anatomical landmarks and require a different amount of training data. The first method is problem-specific and is based on the convex hull of the pelvis. Method two is a more generic approach based on a statistical shape model including the landmarks of interest for every training shape. With our third method we present the most generic approach, where only a small set of training landmarks is required. Those landmarks are transferred to the patient specific geometry based on Mean Value Coordinates (MVCs). The methods work on surfaces of the pelvis that need to be extracted beforehand. We perform this geometry reconstruction with our previously introduced fully automatic segmentation framework for the pelvic bones. With a focus on the accuracy of our novel MVC-based approach, we evaluate and compare our methods on 100 clinical CT datasets, for which gold standard landmarks were defined manually by multiple observers.

***Index Terms*—** Biomedical measurements, Anatomical landmarks, Landmark detection, CT

1. INTRODUCTION

Motivation. The determination of anatomical landmarks is an essential step in a morphological analysis of a solitary bone, but is also a key prerequisite for defining reference systems to assess the relative position of the bones forming a joint. The anterior pelvic plane (APP) is such a reference plane defined by three anatomical landmarks, namely the left and right anterior superior iliac spines (ASIS) and the pubic symphysis [1]. For example the accurate determination of the APP is mandatory for referencing the orientation of the acetabulum [1]. This is a key measure that enables the orthopaedic surgeon to assess important changes in anatomy and the resulting biomechanical conditions due to either disease or as a consequence of surgery [1, 2]. Whilst medical image data provide a valuable source for determining such landmarks, landmark extraction has most often to be performed manually or in a semi-automatic fashion [3]. Unsupervised determination of anatomical landmarks could facilitate an accurate assessment of anatomical changes in individual subjects, and thereby allow for a monitoring of lager patient populations. Thus, less obvious, yet relevant distinctive features may be identified that either predispose to the development of degenerative joint diseases or fuel their further progression [4].

*FP6 IST Project DeSSOS (027252).
†DFG Collaborative Research Center SFB 760.

Related Work. In recent years a number of works focused on the automatic extraction of anatomical point landmarks from medical image data for different application scenarios. With the goal of providing an orthopedic planning tool Ehrhardt et al. [5] proposed a non-rigid registration approach for detecting pelvic landmarks in CT data. They employ an atlas that includes labeled voxels and landmarks. As an initialization for a surface-based registration, Betke et al. [6] introduced a template matching scheme to detect point correspondences in CT images of the lung. Landmark localization based on extremal differential properties like ridges, corners or saddles, was introduced by Wörz and Rohr [7]. The localization is based on the local alignment of 3d parametric deformable models to medical image data. Izard et al. [8] suggest an algorithm for landmark detection based on a probalistic model of image intensities. By learning the image intensities in a set of manually landmarked images, they create a tissue-probability map, which is then aligned to new image data by a likelihood maximization approach. A different method, employing techniques from machine learning, was introduced by Dikmen et al. [9]. They present a three-stage system to roughly locate, verify and finally correct the positions of anatomical landmarks based on previously learned spatial relationships and image features.

Contribution. Most of the studies cited above are limited to a single imaging protocol because they focus on specific *image features*. As a contribution we present and evaluate three fully automatic methods for the extraction of pelvic landmarks based on *geometric reconstructions* of the pelvis. The geometric approach allows for an easy adaptation of our landmark extraction methods to other modalities if the underlying geometric model provides adaptation strategies for other imaging protocols. A fully automatic framework for pelvis segmentation based on a statistical shape model (SSM) as proposed in [10] generates the geometric reconstructions. Our three landmark extraction methods exhibit different degrees of generality and require different amounts of training data. The generic approach, which is based on Mean Value Coordinates (see Sec. 3) requiring only a few landmark training datasets, outperforms a generic approach based on a Statistical Shape Model that includes the landmarks (see Sec. 2) and keeps up with a problem-specific approach that requires no training data. The reconstruction errors on a set of CT data are comparable to inter-observer variability.

2. STATISTICAL SHAPE MODEL

For the generation of a statistical shape model 50 post-operative CT scans with an approximate resolution of $0.9x0.9x5mm^3$ were available from male and female total hip arthroplasty patients (THA). All CT datasets were manually processed by human experts. The pelvic bones were labeled as described in [10]. Additionally, three pelvic landmarks were extracted from each dataset, namely the left

and right anterior superior iliac spines (LASIS and RASIS) and the anterior superior point of the symphysis (SYM), as shown in Figure 1. Note, that those landmarks do not necessarily lie on the bone surface.

Surfaces reconstructed from the manually segmented datasets served as training set for the generation of a statistical shape model of the pelvic bones. Triangular surfaces with corresponding mesh topologies were generated as in [10]. To each of these meshes, the respective anatomical landmarks were added as unlinked points. Principal components analysis on the resulting training set yields a linear model of the form

$$S(\mathrm{b}, T) = T(\overline{v} + \sum_k b_k \mathrm{p}_k) \tag{1}$$

where $\overline{v} \in \mathbb{R}^{3m}$ represents the mean shape, $p_k \in \mathbb{R}^{3m}$ the modes of shape variation (eigenmodes), m being the number of sample points used to discretize the shapes (including the three points that describe the landmarks), $b_k \in \mathbb{R}$ are the shape weights and T is an affine transformation. Any instance of a pelvis comprised in the statistical analysis, i.e. any training surface and three landmarks, can now be represented by such a linear combination.

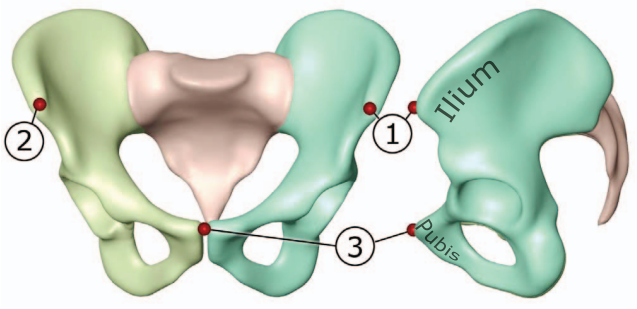

Fig. 1. Average pelvis shape $\overline{v}$ including the anatomical landmarks LASIS (1), RASIS (2) and symphysis (3).

3. MEAN VALUE COORDINATES

Mean Value Coordinates (MVCs) are a generalization of barycentric coordinates. They allow a point in space to be expressed as a convex combination of the four vertices belonging to a surrounding tetrahedron [11]. Ju et al. [12] extended this approach so that instead of a surrounding tetrahedron, an arbitrary closed triangular *control mesh* M can be used to express the position of a point p in space,

$$p = \frac{\sum_j w_j v_j}{\sum_j w_j} \tag{2}$$

with v_j being the vertices of M and w_j the respective *mean value weights*. For details on the computation of w_j see [12]. The control mesh M does not have to be convex, nor does it have to surround p. When changing the control mesh M, i.e. displacing its vertex positions v_j to new positions $\hat{v}_j$, a new point position $\hat{p}$ can be computed by replacing the vertex positions v_j with $\hat{v}_j$ in equation 2. MVCs are well defined inside and outside of M and are suitable for extra- and/or interpolating point positions relatively to the control mesh. We chose to use MVCs for the task of landmark description, although there exist similar methods which might be suitable for this task, like Green Coordinates or Harmonic Coordinates [13, 14].

4. LANDMARK EXTRACTION

As a basis for our landmark extraction we apply an automatic reconstruction method of the pelvic bone from CT data previously published in [10]. With the pelvis SSM as input (see Sec. 2), this method produces accurate geometric reconstructions of the pelvis with consistent mesh topology as follows: After global initialization of the average pelvis shape via the Generalized Hough Transform, the SSM is adapted to the image data, yielding robust segmentations of the CT data (SSM-results). SSM adaptation also produces extrapolated landmark positions. Based on the intermediate SSM-results, graph based optimization is performed to generate accurate segmentations (OPT-results). The landmark positions contained in an SSM-result are not considered in this step, i.e. the resulting accurate segmentation does not contain landmark reconstructions. Note that when segmenting one of the CT volumes used for model generation (THA), we remove the respective training surface from the model. The reconstruction error of the SSM-results in terms of average mean surface distance to gold standard surfaces (AD) is 1 ± 0.3 mm for the 50 THA datasets. The reconstruction error of the OPT-results (AD) is 0.6 ± 0.3 mm for the 50 THA datasets.

4.1. Convex Hull Method

The convex hull method represents a problem-specific approach that works exclusively on the OPT-results and does not require any training landmarks. The method exploits fact that the landmarks of interest can be considered the most prominent anterior points of the pelvic anatomy, defining the anterior pelvic plane (APP). Thus, they are also part of the convex hull of the reconstructed pelvic bones. The method takes the triangulated convex hull H of the OPT-result as input. For extracting the three landmarks, we first identify the triangle t of H whose vertices have the smallest sum of distances from the pubis, the left ilium, and the right ilium (see Figure 1). Note that these regions are defined as *patches* on the OPT-result surface meshes [10]. Thus for each of the three regions, the distance to the closest vertex of a triangle of H can be determined. The triangle t minimizes the sum of the resulting three distances. It is assumed to define the APP. The vertex on the left ilium patch of the OPT-result with minimum distance to the APP is considered the LASIS. The RASIS is defined analogously. For the pubis, two nearest points to the APP are determined, one for each hip bone. The midpoint between the two nearest points is considered the symphysis landmark SYM.

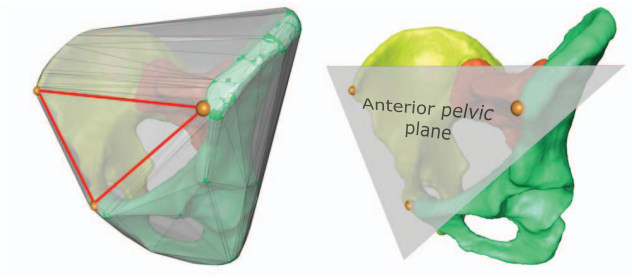

Fig. 2. Convex hull of the pelvic bones (left) including the APP with minimum distance to both ilia and the symphysis (right).

4.2. Statistical Shape Model Method

The SSM approach is a more general method, which defines the landmarks as contained in the SSM-results. It can be applied to any approach using a statistical shape representation in terms of a point distribution model (PDM). Here, extraction of anatomical landmarks by an SSM is performed by including all landmarks of interest from each training shape into the linear model $S(\mathrm{b}, T)$. Now, the adaptation of this model to image data also yields positions of the landmarks. In our approach, these landmark positions are extrapolated by the statistical shape model. No specific image features guide the adaptation of the landmarks. We then compute the Mean Value weights of the landmarks of the SSM-result with respect to the pelvis surface mesh of the SSM-result (see Sec. 3). As any OPT-result is basically a deformed version of the respective SSM-result with the same mesh topology, the Mean Value weights can then be used to transfer the landmark coordinates to the OPT-result.

4.3. Averaged Mean Value Coordinates Method

The SSM approach introduced in Section 4.2 is able to extract anatomical landmarks *on-the-fly* after alignment of the model to image data. However, the landmarks of interest may not be available as training data during the model generation process and an existing SSM is not easily extensible. We propose a generic method suitable to deal with this situation. Our idea is: given an existing SSM of an anatomical structure, the user has to define some reference landmarks for only a small set of training shapes, which are then transferred to any instance of the SSM by means of MVCs.

The linear character of MVCs allows for an easy combination of different weights w_j obtained from multiple control meshes with identical topology. In our case, the Mean Value weights of landmarks of different training shapes may be combined to form average weights that may be used to describe landmark positions with respect to a new, reconstructed pelvis mesh,

$$v = \frac{\sum_j \bar{w}_j p_j}{\sum_j \bar{w}_j} \quad \text{with} \quad \bar{w}_j = \sum_{k \in K} w_j^k \tag{3}$$

with K being the set of training surfaces used as landmark reference, w_j^k the j-th Mean Value weight of the k-th training dataset, and p_j the vertices of the reconstructed pelvis mesh. The averaging of mean value weights is supposed to minimize the influence of outliers from the training set, i.e. reference landmarks which are defined inaccurately. This inter-observer variability is also the reason for not defining the landmarks on only the mean shape $\overline{v}$ and then transferring it to the reconstructed shape. The outcome of such an approach strongly depends on the error performed by the human observer on a single instance of the shape model. By choosing K, which also means to choose its cardinality $|K|$, the user is free to decide whether all reference landmarks may be considered by the method or only a limited set. We chose unweighted averaging as we consider all training shapes equally important: There is no evidence that training shapes far from the mean shape have less influence on the averaged mean value weights.

5. EVALUATION

For the evaluation of our methods 50 CT datasets as used for model generation (THA, see Sec. 2) and an additional 50 CT datasets with a higher resolution of approx. $0.9x0.9x1mm^3$ (Non-THA) were available. For all datasets manually defined landmarks from three human experts were available, namely left and right ASIS as well as the symphysis. We chose one of these reference landmark sets as gold standard whereas the remaining two were used as control group to measure the inter-observer variability of the manual landmark definition process.

As for our three automatic landmark extraction methods, we employed the fully automatic segmentation framework to reconstruct the pelvis from all CT datasets, and applied the methods on the resulting surface meshes, as described in Sec. 4. The Averaged-MVC method (Sec. 4.3) was tested with different numbers $|K|$ of randomly chosen reference datasets. All landmarks were compared to the gold standard in terms of the Euclidean distance. Additionally we compared the orientation of the APP defined by the three extracted landmarks to the gold standard APP. The angle θ between the plane normals serves as an error measure.

6. RESULTS

All results are listed in Table 1. For the THA datasets the manual landmark definition (User2, User3) reveals an inter-observer variability in a range of 5.7 to $6.8mm$ for left and right ASIS landmarks, whereas for the symphisis (SYM) inter-observer variability ranges from 2.5 to $3.4mm$ on average. The resulting APP orientation shows a variability of 0.8° to 1.3°. On the higher resolution datasets the variation for the ASIS landmarks is reduced to values below $3mm$ and a range of 0.5° to 0.7° for the APP orientation.

For the THA datasets, the convex hull method (Hull) and the Averaged-MVC(aMVC) with $|K| \geq 10$ yield results with APP deviation angles θ within the inter-observer variability. The Statistical Shape Model method with successive Mean Value transfer (SSM+MVC) performs worse in terms of θ. An improvement in landmark detection accuracy can be observed on the Non-THA datasets for all methods, if less obvious for SSM+MVC and aMVC. However, the APP deviation angle θ does not improve for any of the automatic methods, but only for the manual cases. As for the automatic methods, the Hull method performs best on the Non-THA datasets in terms of both landmark accuracy and APP deviation angle, closely followed by the aMVC method with $|K| \geq 50$. The SSM+MVC method performs worse again in terms of θ.

7. DISCUSSION

The convex hull method produces results nearly identical to those from human observers on the THA datasets, and only slightly worse on the non-THA datasets. We infer that this method is the most accurate in finding the landmarks of interest. However, it is also the most problem-specific method and cannot be easily extended to cope with additional landmarks that are not defined via the convex hull.

The SSM+MVC and aMVC methods do not produce much better results on the non-THA datasets in terms of landmark accuracy as compared to the THA datasets. This may be attributed to the fact that the training landmarks that are exploited in both methods stem from the low resolution THA datasets, and may therefore exhibit an intrinsic error caused by the large slice distance of 5mm. This is supported by the fact that the Hull method and the manual landmark extractions, working independently from any training landmarks, show better results on higher resolution data.

We assume that the slice resolution of about $5mm$ may be the main reason for the high landmark position errors of all (automatic and manual) cases in the THA datasets. Part of the landmark position error in all automatic methods can be attributed to few cases in which

	THA (n=50 / low-resolution)				**Non-THA** (n=50 / high-resolution)			
	LASIS	RASIS	SYM	θ	LASIS	RASIS	SYM	θ
	mm (std)			*deg* (std)	*mm* (std)			*deg* (std)
User2	**6.2** (3.8)	**6.8** (4.2)	**3.4** (2.1)	**1.3** (1.0)	**2.8** (2.1)	**2.3** (1.5)	**3.2** (1.2)	**0.5** (0.4)
User3	**5.8** (3.2)	**5.7** (2.9)	**2.5** (1.5)	**0.8** (0.6)	**2.9** (2.3)	**2.7** (2.4)	**2.5** (1.2)	**0.7** (0.6)
Hull	**6.9** (4.6)	**6.3** (3.2)	**3.6** (1.8)	**1.0** (0.7)	**3.9** (2.8)	**3.8** (2.5)	**3.6** (1.9)	**1.0** (0.9)
SSM	**7.3** (3.6)	**8.1** (3.9)	**4.3** (2.0)	**2.2** (1.8)	**5.5** (2.7)	**7.1** (3.6)	**4.6** (2.0)	**2.5** (1.4)
SSM+MVC	**6.9** (3.7)	**7.6** (4.0)	**4.0** (2.0)	**1.6** (1.3)	**5.0** (2.7)	**6.8** (3.6)	**4.3** (2.0)	**1.7** (1.2)
aMVC $\lvert K\rvert$=1	**5.8** (3.3)	**6.0** (3.7)	**5.7** (2.3)	**1.9** (1.0)	**6.8** (3.7)	**6.4** (2.9)	**5.1** (2.4)	**1.9** (1.4)
$\lvert K\rvert$=5	**7.3** (4.4)	**6.5** (4.2)	**3.7** (2.0)	**1.4** (0.9)	**6.2** (4.1)	**6.2** (3.1)	**3.5** (1.9)	**1.4** (1.1)
$\lvert K\rvert$=10	**7.8** (4.6)	**7.4** (4.6)	**3.8** (2.1)	**1.3** (0.9)	**6.3** (4.3)	**6.4** (3.4)	**3.7** (1.9)	**1.3** (1.1)
$\lvert K\rvert$=50	**7.3** (4.2)	**7.4** (4.4)	**3.6** (2.1)	**1.3** (1.0)	**6.3** (3.8)	**6.6** (3.2)	**3.8** (2.0)	**1.3** (1.0)

Table 1. Evaluation results (THA and non-THA patients). Average error metrics (bold) and standard deviation (in brackets).

the automatic reconstruction of the pelvis shows inaccuracies in the region of the iliac crest.

Overall our Averaged-MVC method produces results that outperform the SSM+MVC method, and are only slightly worse than the Hull method. Training landmarks are required on only a small subset of the training data for the Averaged-MVC method, whereas the SSM+MVC method requires training landmarks for the complete set of training data at the time of SSM generation.

We rely on the quality of the established correspondences of surface mesh points during SSM generation. Therefore, further studies are required to assess the influence of the landmark location and distance relative to the reference surface on the extraction quality.

8. CONCLUSION

We presented three methods for the extraction of anatomical point landmarks. Each of these methods takes a geometric reconstruction of the pelvis as input. In an evaluation on 100 CT datasets we could show that the results produced by our automatic methods are comparable to those of human experts in terms of accuracy on low resolution datasets. On high resolution data human observers outperform the automatic SSM- and MVC-based approaches for the ASIS landmarks, which may be attributed to the training landmarks which stem from low resolution data. The convex hull approach performs best on high resolution data, however, the average error metrics lie slightly above of the range of inter-user variability.

With our MVC-based approach we could present a generic method which is suitable to perform automatic extraction of anatomical landmarks with a minimum requirement on manual effort, i.e. the definition of sample landmarks on a very small subset of shape instances of a statistical shape model. Together with our automatic reconstruction framework this would allow for efficient landmark extraction by non-experts.

9. REFERENCES

[1] G. E. Lewinnek, J. L. Lewis, R. Tarr, C. L. Compere, and J. R. Zimmerman, "Dislocations after total hip replacement," *J Bone Joint Surg Am*, vol. 60, pp. 217–220, 1978.

[2] M. O. Heller, J. H. Schröder, G. Matziolis, A. Sharenkov, W. R. Taylor, C. Perka, and G. N. Duda, "Musculoskeletal load analysis. a biomechanical explanation for clinical results – and more?," *Der Orthopäde*, vol. 36, pp. 188–194, 2007.

[3] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration.," *Med Image Anal*, vol. 2, no. 1, pp. 1–36, Mar 1998.

[4] J. S. Gregory, J. H. Waarsing, J. Day, H. A. Pols, M. Reijman, J. Weinans, and R. M. Aspden, "Early identification of radiographic osteoarthritis of the hip using an active shape model to quantify changes in bone morphometric features: Can hip shape tell us anything about the progression of osteoarthritis?," *Arthritis & Rheumatism*, vol. 56, pp. 3634–3643, 2007.

[5] J. Ehrhardt, H. Handels, W. Plötz, and S. J. Pöppl, "Atlas-based recognition of anatomical structures and landmarks and the automatic computation of orthopedic parameters.," *Methods Inf Med*, vol. 43, no. 4, pp. 391–397, 2004.

[6] M. Betke, H. Hong, D. Thomas, C. Prince, and J. P. Ko, "Landmark detection in the chest and registration of lung surfaces with an application to nodule registration.," *Med Image Anal*, vol. 7, no. 3, pp. 265–281, Sep 2003.

[7] S. Wörz and K. Rohr, "Localization of anatomical point landmarks in 3D medical images by fitting 3D parametric intensity models," *Med Image Anal*, vol. 10, no. 1, pp. 41–58, 2006.

[8] C. Izard, B. Jedynak, and C. E. L. Stark, "Spline-based probabilistic model for anatomical landmark detection," in *MICCAI (1)*, 2006, pp. 849–856.

[9] M. Dikmen, Y. Zhan, and X. S. Zhou, "Joint detection and localization of multiple anatomical landmarks through learning," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Apr. 2008, vol. 6915 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*.

[10] H. Seim, D. Kainmueller, M. Heller, H. Lamecker, S. Zachow, and H.-C. Hege, "Automatic segmentation of the pelvic bones from ct data based on a statistical shape model," in *Eurographics Workshop on Visual Computing for Biomedicine (VCBM)*, Delft, Netherlands, 2008, pp. 93–100.

[11] M. S. Floater, G. Kós, and M. Reimers, "Mean value coordinates in 3d," *Comp Aid Geom Des*, vol. 22, no. 7, pp. 623–631, October 2005.

[12] T. Ju, S. Schaefer, and J. Warren, "Mean value coordinates for closed triangular meshes," *ACM Trans Graph*, vol. 24, no. 3, pp. 561–566, 2005.

[13] Y. Lipman, D. Levin, and D. Cohen-Or, "Green coordinates," *ACM Trans Graph*, vol. 27, no. 3, pp. 1–10, 2008.

[14] P. Joshi, M. Meyer, T. DeRose, B. Green, and T. Sanocki, "Harmonic coordinates for character articulation," *ACM Trans Graph*, vol. 26, no. 3, pp. 71, 2007.