# AUTOMATED CONSTRUCTION OF 3D STATISTICAL SHAPE MODELS

Tomaž Vrtovec[1], Dejan Tomaževič[1], Boštjan Likar[1], Ludvik Travnik[2] and Franjo Pernuš[1]

[1]Department of Electrical Engineering, University of Ljubljana, Tržaška 25, 1000 Ljubljana, Slovenia, [2]Department of Orthopaedic Surgery, Clinical Center, Zaloška 9, 1000 Ljubljana, Slovenia
e-mail: {tomaz.vrtovec, dejan.tomazevic, bostjan.likar, franjo.pernus}@fe.uni-lj.si
*(Accepted February 5, 2004)*

## ABSTRACT

Automated segmentation of medical images is a difficult task because of the complexity of anatomic structures, inter-patient variability, and imperfect image acquisition. Prior knowledge, in the form of point-based statistical shape models (point distribution models) of a structure of interest can greatly assist segmentation to robustly find the structure in a patient's image. Point distribution models are obtained through sets of corresponding landmarks lying on surfaces of training structures. The key to the automated construction of a three-dimensional (3D) statistical shape model is the identification of corresponding landmarks on training shapes, which is a challenging task. This paper presents a novel method for automated construction of 3D point distribution models. Corresponding surface points are obtained by two main steps: 1) volumes of interest (VOI), each containing one training structure, are manually defined, a reference structure is manually extracted from one training VOI and its surface is established and represented by a set of (reference) points, 2) reference landmarks are propagated to other training VOIs by transformations that are obtained by hierarchical elastic registration between the reference and each of the remaining training VOIs. We illustrate our approach using computed tomography data of the lumbar vertebra.

Keywords: lumbar vertebrae, non-rigid image registration, principal component analysis, statistical shape models.

## INTRODUCTION

Segmentation of medical images, which is the task of partitioning the image data into contiguous regions representing individual anatomic structures is a prerequisite for further investigations in many computer-assisted medical applications, e.g. diagnosis, therapy planning and evaluation, simulation and image guided surgery. Automated segmentation of medical images is a very difficult task because of the complexity of anatomic structures, inter-patient variability, and imperfect image acquisition. Prior knowledge, in the form of statistical models of shape variability or point distribution models (PDMs), of a structure of interest can greatly assist automated segmentation to robustly and accurately find the structure in a patient's image. A PDM describes shape variability across a set of training structures through variations in positions of a set of corresponding landmarks (Cootes *et al.*, 1995). The 'modes of shape variation' are obtained through principal component analysis (PCA) of the covariance matrix of landmark coordinates. Since their introduction, PDMs have

been applied to various anatomical structures, like the vertebrae (Lorenz and Krahnstover, 2000; Kaus *et al.*, 2003), femur (Fleute *et al.*, 1999), liver (Lamecker *et al.*, 2002), lung (Li and Reinhardt, 2001), heart (Frangi *et al.*, 2002) and brain structures (Duta and Sonka, 1997).

In building statistical models, a set of corresponding landmarks defined over the set of training shapes is required. Manual determination of point correspondences is a tedious, time-consuming, and user dependent task. This is particularly true for 3D structures, where the amount of landmarks required to describe the shape increases dramatically in comparison to 2D structures. On the other hand, automated identification of corresponding landmarks on training shapes is a demanding task. The goal of the present work is to automate the determination of landmarks and correspondences between them. Several authors have already proposed automated techniques to find point correspondences for building PDMs. The methods proposed by Fleute et al. (1999), Lamecker *et al.* (2002), Kaus *et al.* (2003), Brett and Taylor (2000), and Frangi *et al.* (2002) require segmented

training objects. One segmented anatomical structure is selected as the reference shape and its surface is represented by a mesh (triangulation) consisting of vertices and triangles. From this point on, the methods differ in the way the vertices of the mesh, which are the landmarks or points, representing the reference structure are propagated to surfaces of other training structures. Fleute *et al*. (1999) perform the alignment and matching process using a multi-resolution approach based on octree splines. Lamecker *et al.* (2002) decompose the surfaces into patches and map each patch onto a disk. Two corresponding disks give a one-to-one mapping from one patch to another, resulting in transferring one triangulation onto the other surface. For each of the remaining learning shapes, Kaus *et al.* (2003) first estimate a preliminary pose and scaling that approximately aligns the template mesh to the learning shape and then adapt the template mesh to the learning shape using a deformable model based approach. The method of Brett and Taylor (2000) uses the symmetric version of the iterative closest point (ICP) algorithm to establish correspondences between sparse vertices on the reference and each of the learning shapes. Frangi *et al.* (2002) register the reference and a training shape by the algorithm of Rueckert *et al.* (1999) which manipulates a shape by embedding it into a subsequently refined volumetric mesh, which defines a continuous deformation field through a set of B-spline functions. The corresponding optimal deformation is obtained by maximizing a voxel similarity measure on the basis of the corresponding labels.

The above methods require segmentation of all training structures to derive a realistic model of shape variations. However, accurate segmentation of complex anatomical structures is not an easy task and is for the purpose of model building usually performed manually. In this paper a novel method for automated construction of 3D PDMs is proposed, which requires the segmentation of only one training structure. Corresponding surface points are obtained by two main steps: 1) volumes of interest (VOIs), each containing one training structure, are manually defined, a reference structure is manually extracted from one VOI and its surface is established and represented by a set of (reference) points, 2) reference landmarks are propagated to other training VOIs by transformations that are obtained by hierarchical elastic registration between the reference and each of the other training VOIs. We illustrate our approach using computed tomography data of the lumbar vertebra.

## AUTOMATED LANDMARKING

The input to our method are $n$ manually defined VOIs, each containing one training structure. For each of the training structures (shapes) we seek a set $x_i$, $i = 1, 2, \dots, n$ of $m$ landmarks

$$x_i = \{ p_j = ( p_{1,j}, p_{2,j}, p_{3,j} ), j = 1, 2, \dots, m\}, \quad (1)$$

where the $j$th, $j = 1, 2, \dots, m$ landmark corresponds to the same surface location from one shape to another. In medical images a landmark is usually an anatomically characteristic point that can be uniquely identified on a set of the same anatomic structures. Because generally too few anatomic landmarks can be uniquely identified to accurately describe the shape of a 3D structure, we use pseudo-landmarks – vertices of the triangulation of a reference structure's surface. The dense triangulation of the isosurfaces can be further decimated to reduce the amount of vertices. We propagate the reference point set $x_{ref} = x_1$, defined by the reference training structure, to other training structures by non-rigidly registering the reference VOI to each of the remaining training VOIs. Fig. 1 illustrates our automated landmarking framework. The following sections describe in more detail the main steps of establishing point correspondences.

### LANDMARKING THE REFERENCE STRUCTURE

A typical 3D (set of sections) computed tomography (CT) VOI containing a lumbar vertebra is selected as the reference VOI (image). On this image, we manually segment the vertebra to obtain its shape. From the segmented vertebra, we build its surface model by the isosurface method. An isosurface is a spatial function that connects all the points in space that hold the same function value, or isovalue. The result is the shape model of the reference vertebra in the form of a triangle mesh, consisting of $m$ vertices (Fig. 2). The vertices serve as the reference set of points (landmarks) $x_{ref} = x_1$. In the next steps, $x_1$ has to be propagated to other training VOIs (structures) to obtain the corresponding point sets $x_i$, $i = 2, 3, \dots, n$.
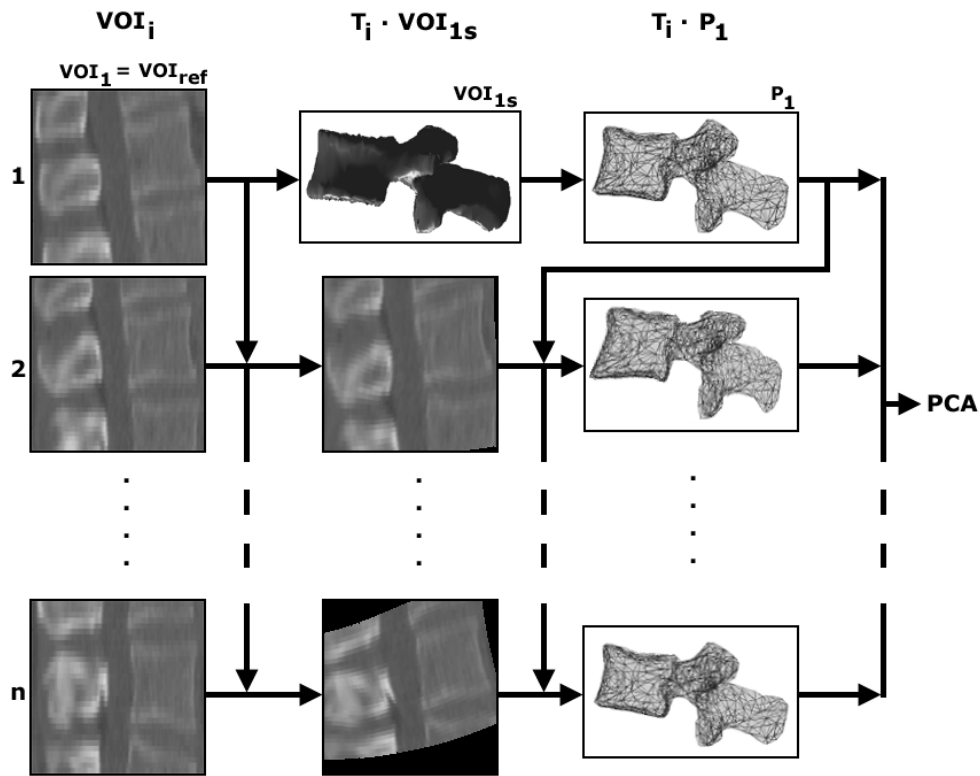
Fig. 1. *The automated landmarking framework. A set of reference landmarks are established from the reference VOI (top row). By registering the other training VOIs (left column) to the reference VOI, transformations $T_i$ are determined (middle column) and used to propagate the set of reference landmarks to other training VOIs (right column). Finally, the PDM is obtained by performing a statistical analysis on obtained sets of landmarks.*
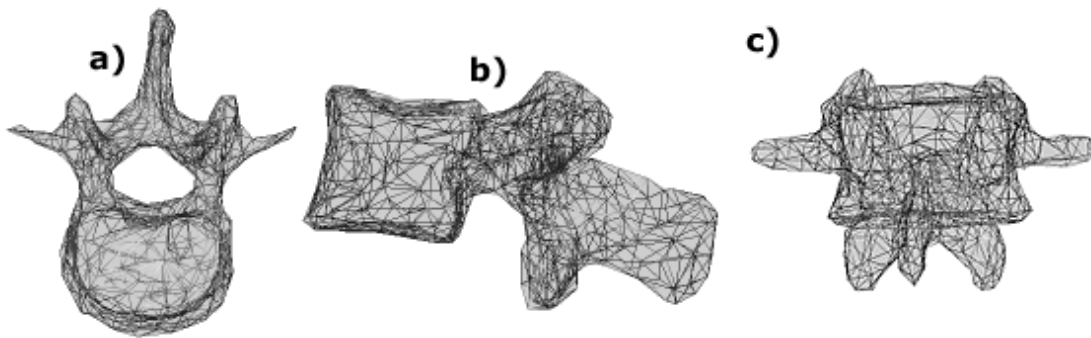


Fig. 2. *Shape model of the reference vertebra: a) axial, b) sagittal and c) coronal views.*

## PROPAGATION OF REFERENCE POINTS

Propagation requires the determination of *n-1* transformations $T_i$, $i = 2, 3, \ldots , n$ , by which the points in the reference point set $x_1$ are transformed to approximately the same locations on surfaces of remaining training structures

$$x_i = T_i \cdot x_1; \, i = 2, 3, \ldots , n. \qquad (2)$$

The transformations $T_i$, $i = 2, 3, \ldots , n$ are obtained by non-rigidly registering the reference VOI with each of the other training VOIs. The hierarchical approach to elastic registration (HER) is applied (Likar and Pernuš, 1999; 2001). By this method, the images to be registered are progressively subdivided, locally registered, and elastically interpolated. Fig. 3 illustrates the 3D registration procedure. Each pair of subimages is registered by finding those parameters of the affine transformation that maximise the similarity measure, which was the normalised mutual information $NMI(VOI_{ref}, T \cdot VOI)$ (Studholme *et al.*, 1999)

$$NMI\left(VOI_{ref}, T \cdot VOI\right) = \frac{H\left(VOI_{ref}\right) + H(T \cdot VOI)}{H\left(VOI_{ref}, T \cdot VOI\right)}, \qquad (3)$$

where $H(VOI_{ref})$ and $H(T \cdot VOI)$ are Shannon entropies of $VOI_{ref}$ and transformed volume of interest $T \cdot VOI$, respectively and $H(VOI_{ref}, T \cdot VOI)$ is their joint entropy.

The optimal affine transformation $T_{opt}$ brings $VOI$ into $T_{opt} \cdot VOI$, containing as much information as possible about $VOI_{ref}$

$$T_{opt} = \underset{T}{arg\ max} NMI\left(VOI_{ref}, T \cdot VOI\right). \qquad (4)$$

Once all VOIs are locally registered at a certain level, the globally consistent registration is achieved by using the elastic thin-plate spline interpolation $T_{TPS}$ on registered centers $c_i = (x_i, y_i, z_i)$ of subimages

$$T_{TPS}(p) = A_T \cdot p + t + \sum_{i=1}^{n} w_i \cdot U\left(\| c_i - p \|\right), \ (5)$$

where $p = (x, y, z)$ are the coordinates of points in $VOI$, $w_i$ are weights of the radial interpolation function $U(r) = |r|$, $A_T$ is the affine transformation matrix and $t$ is the translation vector.

The result of HER are the transformations $T_i, i = 2, 3, ..., n$, required to propagate the reference point set to other training structures (Eq. 2).

## POINT DISTRIBUTION MODEL

Point distribution models describe shape variations using statistics of positions of corresponding landmarks defined on training shapes (Cootes *et al.*, 1995). By aligning $n$ shape samples, consisting of $m$ landmarks, and applying the principal component analysis on the sample distribution, any sample $x$ of the distribution can be expressed as a linear combination of eigenvectors $q$ (Eq. 10). In our case, the samples $x_i, i = 1, 2, ..., n$ are composed of the 3D coordinates of the landmarks (see Eq. 1), with a one-to-one correspondence between the vector elements of a given index ($k^{th}$ element of vector $x_i$ corresponds to the $k^{th}$ element of vector $x_j$) due to point correspondences. In a 3D model $n-1$ eigenvectors generally form the principal basis function, because the number of training shapes is usually smaller than the number of components of sample $x$. The eigenvalues, to which eigenvectors correspond, provide a measure of the compactness of the distribution along each axis defined by the eigenvectors. Prior to PCA, to maximize the specificity (compactness) of the model, the shapes, i.e. vectors $x_i$ are aligned by translation, rotation, and scaling using the Procrustes Analysis (Goodall, 1991; Dryden and Mardia, 1998). The eigenvalues and eigenvectors are obtained as follows. First, the mean shape vector $\bar{x}$ and the
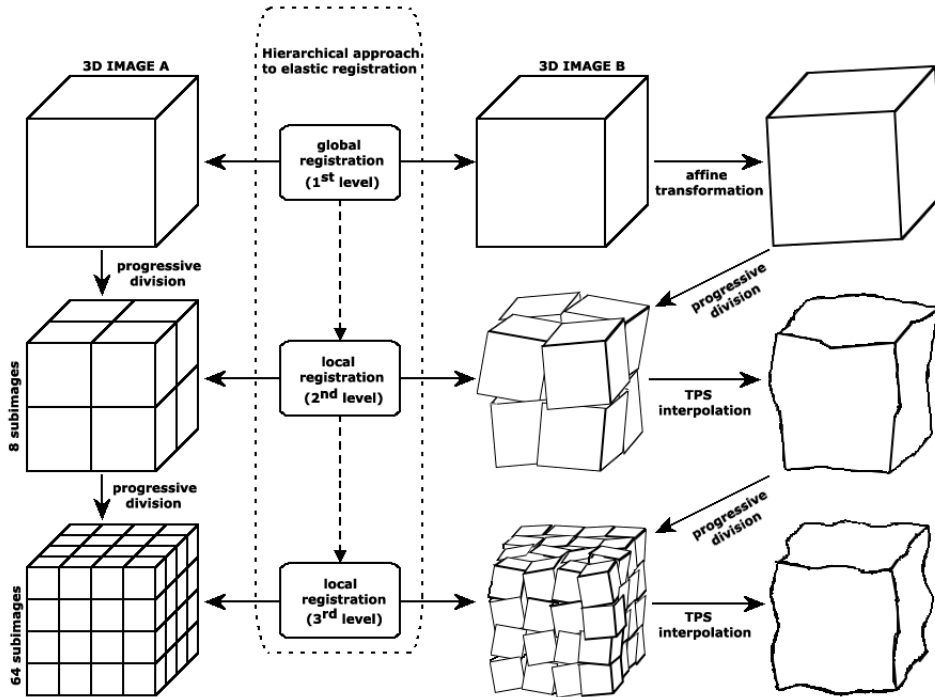


Fig. 3. *The hierarchical approach to elastic 3D registration.*

covariance matrix $C$ that describes the correlation between vector elements are defined

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i , \qquad (6)$$

$$C = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})\cdot(x_i - \bar{x})^T . \qquad (7)$$

By building a matrix $X$ of centered shape vectors $x_i$ - $\bar{x}$, the covariance matrix is simply determined as

$$C = X \cdot X^T . \qquad (8)$$

Performing PCA (Dryden and Mardia, 1998), also known as Karhunen-Loeve transformation, on the covariance matrix $C$ results in $3m$ eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{3m} \geq 0$ and $3m$ corresponding eigenvectors $q_k$; $k = 1,2,...,3m$. The eigenvector $q_1$ that corresponds to the largest eigenvalue $\lambda_1$ contains the largest portion of shape variations. If the number of training set vectors $n$ is smaller than the shape vector dimension, the decomposition is simplified into finding $n$ eigenvalues of the implicit covariance matrix $\tilde{C} = X^T \cdot X$. The first $(n-1)$ eigenvalues and eigenvectors of the matrix $C$ are determined as

$$\lambda_i = \tilde{\lambda}_i; \ q_i = \frac{X\cdot\tilde{q}_i}{|X\cdot\tilde{q}_i|}; \ i = 1,2,...,(n-1) , \qquad (9)$$

where $\tilde{\lambda}_i$ are eigenvalues and $\tilde{q}_i$ are eigenvectors of the implicit covariance matrix $\tilde{C}$. By using these eigenvalues and eigenvectors, it is not only possible to reconstruct original shape vectors, but also to generate new shape vectors (eigen shapes) $x_i'$

$$x_i' = \bar{x} + w_i \cdot q_i , \qquad (10)$$

where $w_i$ are weight vectors. By applying limits to the variation of weight vectors, for instance $|w_i| \leq \pm 3\sqrt{\lambda_i}$, it can be ensured that a generated shape is similar to those contained in the training set.

## EXPERIMENTS AND RESULTS

The image database consisted of 10 3D CT images of the lumbar spine (L1-L5) of the same voxel sizes (0.488×0.488×2 mm$^3$) of 10 individuals, which underwent scanning due to abdominal aortic aneurysm. Fig. 4 shows axial, sagital, and coronal views of some of the CT images. Twenty five VOIs (sizes from 242×187×31 to 261×217×41 pixels) were manually defined, each containing one whole vertebra (4 L1, 7

L2, 6 L3, 5 L4 and 3 L5). These VOIs served as the training set ($n = 25$). Fig. 5 shows the axial, sagittal, and coronal views of some of the VOIs. A VOI containing vertebra L2 was selected as a reference VOI, manually segmented and its surface triangulated by using the isosurface routines in mathematical software (Matlab, MathWorks). Fig. 2 shows the segmented reference vertebra with the overlayed triangular mesh. The original number of vertices was 120000 and was decimated to 12000 vertices, which represented the reference landmarks. We have chosen the 10% decimation to speed up the computation, while the number of landmarks remained big enough to reliably describe shape characteristics. The landmarks were propagated to other training VOIs as described in Section 2. The number of levels in the hierarchical approach to elastic registration was 3.

Table 1 illustrates the portions of shape variability (the individual and relative cumulative shape variance) within the training sample captured by the first 5 eigenmodes.

Table 1. *Individual and relative cumulative shape variance captured by the first 5 eigenmodes.*

| | | eigenvalue $\lambda_i$ | |
| | | % of capture | |
| $i$ | value | individual | cumulative |
|---|---|---|---|
| 1 | 12.05 | 52.03 | 52.03 |
| 2 | 3.63 | 15.70 | 67.73 |
| 3 | 3.03 | 13.11 | 80.84 |
| 4 | 2.03 | 8.79 | 89.63 |
| 5 | 1.46 | 6.33 | 95.96 |

The first eigen shape $x_1'$ is determined by the largest eigenvalue $\lambda_1 = 12.05$ and the corresponding eigenvector $q_1$, and is equal to

$$x_1' = \bar{x} + w_1 \cdot q_1 = \bar{x} \pm k\sqrt{\lambda_1}q_1 \ ; \ k = 0,1,2 . \ (11)$$

The use of PCA as a statistical description provides a compact way to model variations of the structures. The first eigen shape captures about 52% of the variations of the structures present in the image database. By varying the parameter k (Eq. 11), the mean shape is transformed in the direction that is described by the most significant variation mode, i.e. first eigen mode, captured by the PCA. The obtained new shape instances retain as much statistical information of the structures in the training set as it is included in this most significant mode. The resulting propagation and contraction of the vertebra, dictated by the first eigen mode, is clearly visible in each view (Fig. 6).

Further eigen shapes are determined similarly to the first one, where the second eigen shape $x_2'$ ($\lambda_2 = 3.63$) captures about *16%* of variation (Fig. 7) and the

third eignen shape $x'_3$ ($\lambda_3 = 3.03$) about *13%* of variation (Fig. 8). Eigen shapes that follow do not capture significant variations as their eigenvalues are too small in comparison to the largest eigenvalue.
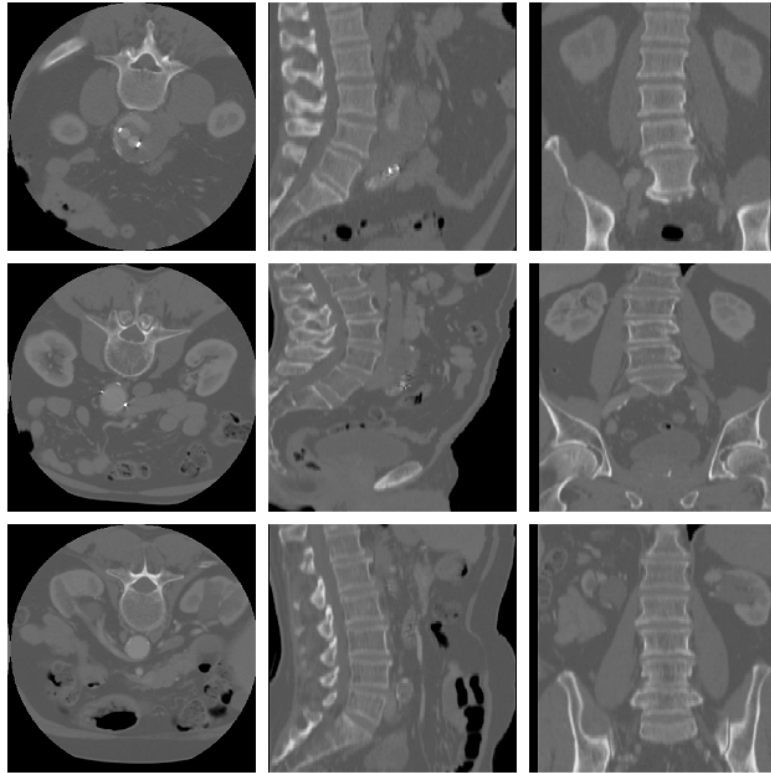


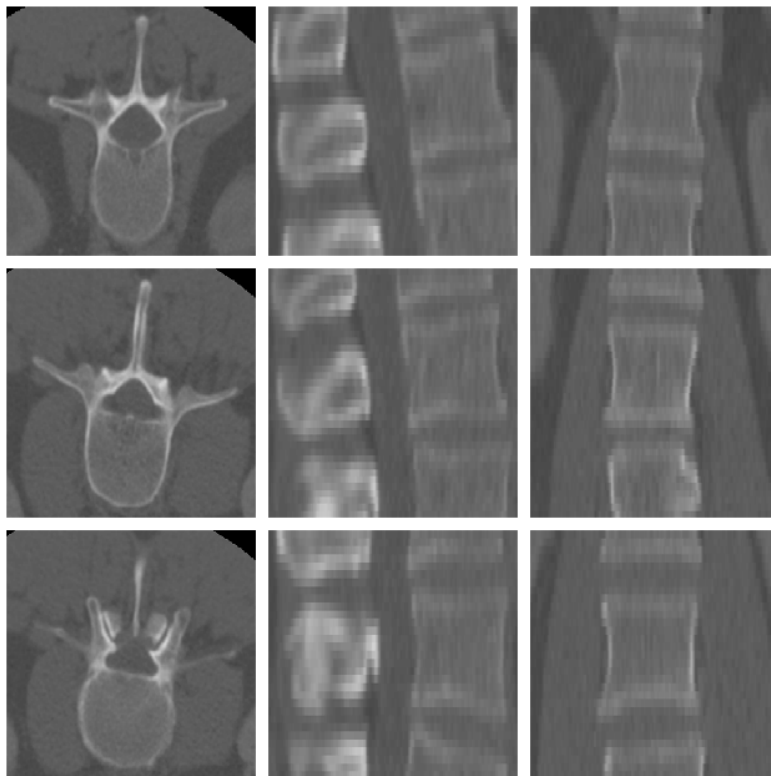Fig. 4. *Axial (left), sagittal (middle), and coronal (right) views of three training images.*



Fig. 5. *Axial (left), sagittal (middle), and coronal (right) views of three VOIs containing a single vertebra.*
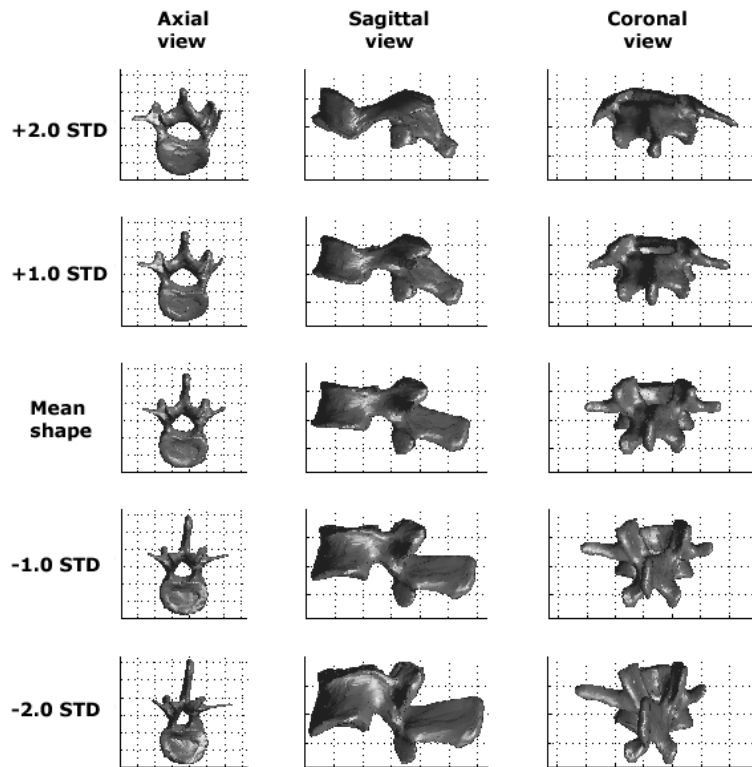
Fig. 6. *The effect of varying the first eigen shape by varying the parameter k in Eq. 11.*
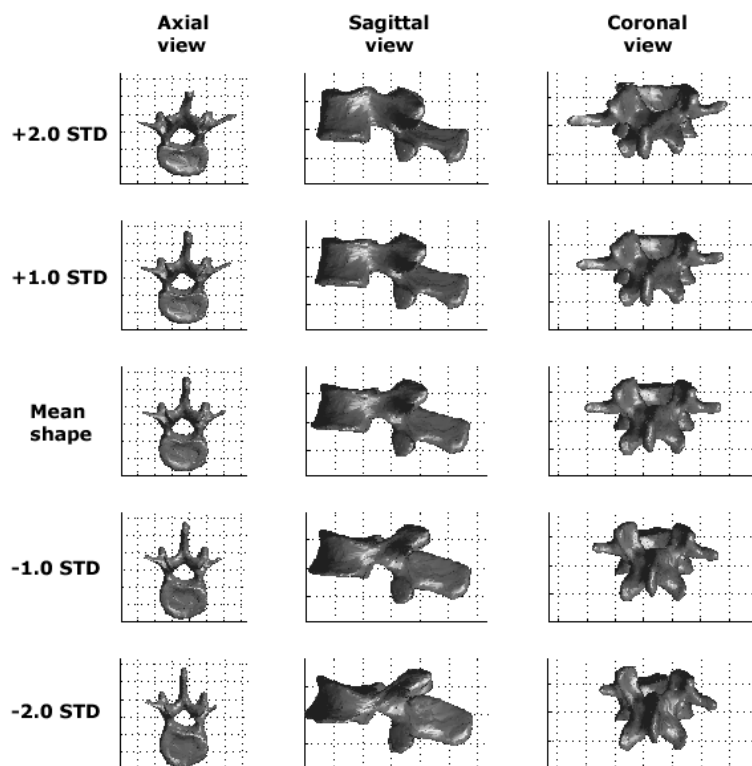
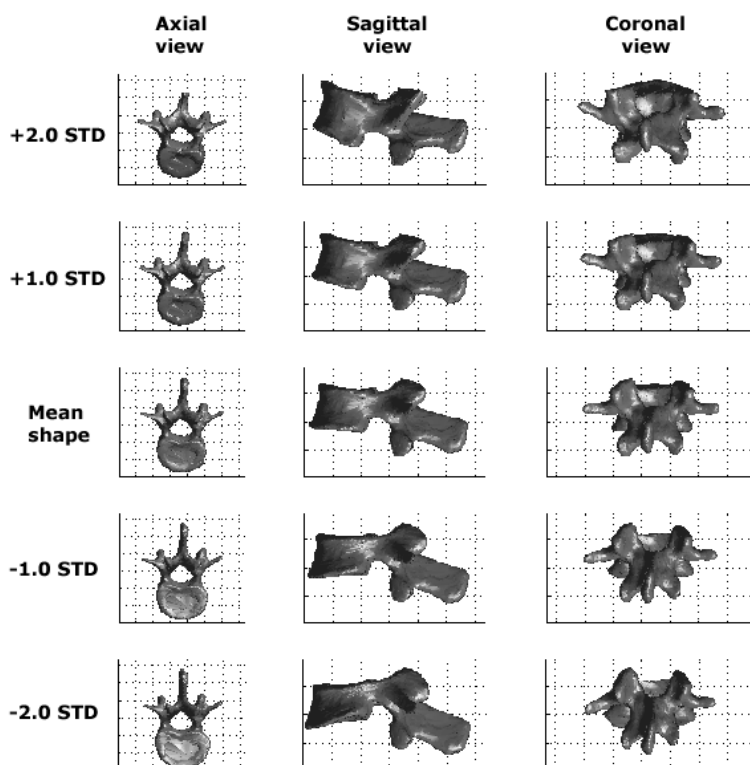Fig. 7. *The effect of varying the second eigen shape.*

Fig. 8. *The effect of varying the third eigen shape.*

## DISCUSSION AND CONCLUSION

In this paper, we have presented a method to automatically construct 3D PDMs from unsegmented 3D images. Corresponding landmarks on surfaces of training structures are established automatically by transformations of reference landmarks defined by a triangular mesh of the segmented reference structure. Transformations are obtained by hierarchical elastic registration between the reference and each of the other training VOIs. The results presented in the previous section are of preliminary nature. For a statistical shape model that would describe all significant variations of vertebra shapes, a larger database would be needed. Especially interesting would be the issue, whether the eigenvalues converge for an increasing number of images. Consequently, the sufficient number of images in the training set could be determined. Moreover, because vertebrae L1-L5 show quite large shape differences, not all vertebra L1-L5 should be modelled by a single model, but probably 5 training databases, one for each of L1-L5 vertebra, should be formed to derive statistical shape models for individual vertebrae L1-L5. Despite the preliminary nature of the results, the described method of generating the statistical shape deformable model is promising. The model well describes the characteristics of the training set of shapes. The deformations included in eigenvalues and corresponding eigenvectors are smooth. Therefore, the model is general for the training set in question, and is at the same time specific enough for the shape class of lumbar vertebrae.

A key step in building a point distribution model involves establishing correct correspondences between points, describing shape boundaries, over an adequate number of training images. If the correspondences are not correct an inefficient parameterisation of shape will be determined. What is even worse, false correspondences will hamper the extraction of new knowledge related to diseases and normal development if these are based on shape analysis. Unfortunately, there is no generally accepted definition for anatomically meaningful correspondences. It is thus difficult to judge the correctness of an established correspondence. The question that arises is how the incorrect correspondences might affect the following PCA analysis for which a strict homology of landmarks is required (Bookstein, 1997). In the proposed method, the correspondences of landmarks are obtained automatically via non-rigid registration. However, if we assume that the inaccuracies in correspondences are distributed randomly over the shapes in the training set, which is also the case when the correspondences are defined manually, the effect on the subsequent PCA shape analysis could be neglected.

These inaccuracies are captured by least significant eigen modes as they are poorly correlated. The problem would be if non-rigid registration would induce systematic errors that would be captured by more significant eigen modes, which is however not very likely for non-rigid registration methods. Nevertheless, the problem should not be underestimated and calls for extensive evaluation, which is far from trivial and beyond the scope of this paper. Very recently, however, Styner *et al.* (2003) have proposed four different measures to measure correctness of correspondences. They have analysed the direct correspondence via manually selected landmarks as well as the properties of the shape model implied by the correspondences, regarding compactness, generalisation, and specificity. We have judged the correctness of correspondences indirectly, by visually assessing the quality of non-rigid registration. The more accurate the registration is, the better the generated model. In our case, the registration results were estimated to be of a medium accuracy. Reasons for this can be found in the non-rigid registration method itself and in possible abnormal – disease or age affected vertebrae shapes that were not noticed while selecting the training image database. The proposed method avoids segmentation of each training image in the database. Such an approach reduces manual interventions as only the reference structure has to be manually segmented. However, the consequence of registering original shapes is that we model shape variations less well. Thus, there is a trade off between goodness of correspondences and time needed to obtain the training structures and point correspondences.

The proposed method can be used for constructing models of arbitrary structures and not only vertebrae. The statistical shape deformable model can then be applied in a relatively new approach to image segmentation, where the model is put in the vicinity of the structure in the image undergoing the segmentation process and then deformed in such a way that it fits the structure as good as possible. For instance, Davatzikos *et al.* (2002) have used a deformable shape model of the spine to automatically find the shape transformation that places patient data into a stereotaxic space. The main application of their method is in lesion-deficit analysis for determining associations between structural damage and clinical symptoms. Benameur *et al.* (2003) registered the 3D statistical shape model expressing variations of the pathological deformations observed on a representative scoliotic vertebra population to two conventional radiographic views (posteroanterior and lateral) to provide knowledge of the 3D structure of the whole scoliotic spine. Automatic image segmentation has also a promising role in image-guided surgery. For instance, Fleute *et al.* (1999) have used statistical shape models to assist anterior cruciate ligament reconstruction.

Further work will include the generation of a shape model by using a larger image database. As individual vertebrae L1-L5 are planned to be modeled, the models will be more specific, increasing the applicability of the model to segmentation and registration procedures.

## REFERENCES

Benameur S, Mignotte M, Parent S, Labelle H, Skalli W, de Guise J (2003). 3D/2D registration and segmentation of scoliotic vertebrae using statistical models. Computerized Medical Imaging and Graphics 27:321-37.

Brett AD, Taylor CJ (2000). A method of automated landmark generation for automated 3D PDM construction. Image and Vision Computing 18:739-48.

Cootes TF, Taylor CJ, Cooper DH, Graham J (1995). Active shape models – Their training and application. Comput. Vision Image Understand 61:38-59.

Davatzikos C, Liu D, Shen D, Herskovits EH (2002). Spatial normalization of spine MR images for statistical correlation of lesions with clinical symptoms. Radiology 224:919-26.

Dryden I, Mardia K (1998). Statistical Shape Analysis. Wiley Series in Probability and Statistics.

Duta N, Sonka M (1997). Segmentation and interpretation of MR brain images: An improved active shape model. IEEE Trans Med Imag 17:1049-67.

Fleute M, Lavallee S, Julliard R (1999). Incorporating a statistically based shape model into a system for computer-assisted anterior cruciate ligament surgery. Medical Image Analysis 3:209-22.

Frangi AF, Rueckert D, Schnabel JA, Niessen WJ (2002). Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modelling. IEEE Trans Med Imag 21:1151-66.

Goodall C (1991). Procrustes methods in the statistical analysis of shape. J Royal Statistical Society B 53:285-339.

Kaus MR, Pekar V, Lorenz C, Truyen R, Lobregt S, Weese J (2003). Automated 3-D PDM construction from segmented images using deformable models. IEEE Trans Med Imag 22:1005-13.

Lamecker H, Lange T, Seebass M (2002). A Statistical Shape Model for the Liver. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2002, Dohi T and Kikinis R (eds.). LNCS 2489:421-7.

Li B, Reinhardt JM (2001). Automatic generation of 3-D shape models and their application to tomographic image segmentation. In Proc SPIE Conf Medical

Imaging, vol. 4322, pages 311-322, San Diego, CA, 17-22 Feb. 2001.

Likar B, Pernuš F (1999). Registration of serial transverse sections of muscle fibers. Cytometry 37:93-106.

Likar B, Pernuš F (2001). A hierarchical approach to elastic registration based on mutual information. Image and Vision Computing 19:33-44.

Lorenz C, Krahnstover N (2000). Generation of point-based 3D statistical shape models for anatomical objects. Comput Vision Image Understand 77: 175-91.

Rueckert D, Sonoda LI, Hayes C, Hill DLG, Leach MO, Hawkes DJ (1999). Non-rigid registration using free-form deformations: Application to breast MR images. IEEE Trans Med Imag 18:712-21.

Studholme C, Hill DLG, Hawkes DJ (1999). An overlap invariant entropy measure of 3D medical image alignment. Pattern Recogn 32:71-86.

Styner MA, Rajamani KT, Nolte L-P, Zsemlye G, Szekely G, Taylor CJ, Davies RH (2003). Evaluation of 3D correspondence methods for model building. Information Processing in Medical Imaging 2003, Ambleside UK.