Preprinted from Frangi, A. F., Radeva, P., Santos, A., and Hernandez, M., editors, Proceedings of the Third International Workshop on Functional Imaging and Modeling of the Heart (FIMH 2005), Lecture Notes in Computer Science, volume 3504, pages 92-101.

Artificial enlargement of a training set for statistical shape models: Application to cardiac images

J. Lötjönen[1], K. Antila[1], E. Lamminmäki[1], J. Koikkalainen[2], M. Lilja[2], and T. Cootes[3]

[1] VTT Information Technology, P.O.B. 1206, FIN-33101 Tampere, Finland
{Jyrki.Lotjonen@vtt.fi }
[2] Laboratory of Biomedical Engineering, Helsinki University of Technology, P.O.B. 2200, FIN-02015 HUT, Finland
[3] Division of Imaging Science and Biomedical Engineering, University of Manchester, U.K.

Abstract. Different methods were evaluated to enlarge artificially a training set which is used to build a statistical shape model. In this work, the shape model was built from MR data of 25 subjects and it consisted of ventricles, atria and epicardium. The method adding smooth non-rigid deformations to original training set examples produced the best results. The results indicated also that artificial deformation modes model better an unseen object than an equal number of standard PCA modes generated from original data.

1 Introduction

Segmentation is known to be one of the most difficult problems in image analysis. Several reasons explain the difficulty, such as noise in images, image inhomogenities, partial volume effect, complex and cluttered scenes and low visibility of edges between objects. Deformable model-based methods provide one approach to overcome partially the problem. In these methods, an *a priori* model is non-rigidly registered to the object of interest in the image by optimizing a cost function. To find the optimal non-rigid transformation is an ill-posed problem; hence some form of constraints are required. One possibility is to use physical-based models, such as viscous fluid or elastic models [1, 2]. Statistical shape models is another option. In these methods, an *a priori* model is allowed to deform only in a way consistent with the information captured from a training set.

The most popular approach for modeling the shape changes is the point distribution model, also referred to as active shape model (ASM) [3]. It defines a mean model and its typical deformation modes on the basis of a training set using principal component analysis (PCA). The deformation modes are the eigenvectors of the covariance matrix determined for corresponding points in different examples of the training set.

Statistical shape models suffer from two commonly known problems especially in medical applications. First, building a statistical shape model is laborious as the point correspondences need to be defined between the training set examples. However, automatic procedures have been recently proposed to overcome this problem [4–6]. Second, as the building process is time consuming and enough data are not always available, only a small set of examples are often used to construct the model. Since the maximum

number of deformation modes can not exceed the number of examples in the training set minus one, the eigenvectors obtained span only a small subspace which can not represent the full range of shape variation present in real medical objects. For example, $10 - 30$ subjects have been used to construct a 3D statistical shape model of the brain and heart [5, 7–9]. As reported in [9], the size of the training set was considered to be the most important reason for a relatively high segmentation error. This work concentrates on the modeling of the heart.

The segmentation accuracy depends on the ability 1) of the model to represent an unseen object, and 2) of the segmentation algorithm to define correctly the model parameters. This work concentrates on the first point. The objectives of this work are two-fold:

- To define an appropriate method for enlarging the training set artificially and efficiently.
- To estimate the relation between the size of the training set and the ability of the model to represent an unseen object, as applied to cardiac data.

The latter objective is closely related to commonly known bias-variance trade-off. In other words, if the number of the deformation modes is too small, the model is overconstrained. However, choosing too many deformation modes based on a large training set leads to overfitting.

2 Methods

2.1 Statistical Shape Models

In statistical shape models, new examples of the shape, $\mathbf{x} = [x_1, \ldots, x_n]^T$, that are specific to the studied object, are generated using a linear combination

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{\Phi}\mathbf{b}, \tag{1}$$

where $\bar{\mathbf{x}} = [\bar{x}_1, \ldots, \bar{x}_n]^T$ is a reference shape, typically a mean shape constructed from a training set, $\mathbf{\Phi} = [\phi_1, \ldots, \phi_m]$ is a matrix consisting of the modes of shape variation, ϕ_i, and $\mathbf{b} = [b_1, \ldots, b_m]^T$ is a weight vector.

In ASM [3], the object is represented by a point set. The training set is first affinely aligned, and the mean shape is calculated:

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i, \tag{2}$$

where N is the size of the training set. Next, PCA is applied to the variations of the training set, i.e., the eigenvectors and eigenvalues of the covariance matrix

$$\mathbf{\Sigma} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}_i - \bar{\mathbf{x}}\right)\left(\mathbf{x}_i - \bar{\mathbf{x}}\right)^T \tag{3}$$

are calculated. The eigenvectors of the covariance matrix describe the ways in which the shapes vary, and the corresponding eigenvalues explain the variance of the data projected onto each eigenvector.

If the number of points in a training set example is denoted by P and the dimensionality by D ($D = 2$ in 2D and $D = 3$ in 3D), the maximum number of deformation modes is $min(N - 1, D \cdot P)$.

The model instance $\mathbf{x}'$ representing an unseen object $\mathbf{x}$ is

$$\mathbf{x}' = \bar{\mathbf{x}} + \mathbf{\Phi}\mathbf{b}', \tag{4}$$

where the weights $\mathbf{b}'$ are computed from

$$\mathbf{b}' = \mathbf{\Phi}^{-1}(\mathbf{x} - \bar{\mathbf{x}}). \tag{5}$$

In this work, we model five objects: 2 atria, 2 ventricles and epicardium. Contours of these objects are catenated to one vector, i.e. $\mathbf{x}_i = (x_{i1}, y_{i1}, x_{i2}, .., y_{iP})$ in 2D, i.e. each mode contained deformations for all objects.

2.2 Materials

Our dataset consisted of cardiac short- and long-axis magnetic resonance images acquired from 25 healthy subjects. The mean shape and its variation were modeled as described in detail in [9]. The procedure is shortly summarized.

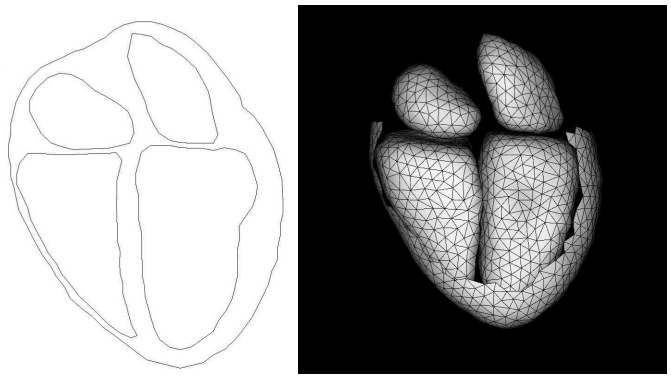
The atria, ventricles and epicardium were manually segmented by fitting a triangulated surface model simultaneously to the short- and long-axis images. Thereafter, one subject was considered as a reference volume to which all other subjects were aligned using translation, rotation and isotropic scaling. The normalized mutual information (NMI) was used as a similarity measure. Segmented volumes, where each object was represented by one gray-scale value, were used in registration. Next, the reference volume was non-rigidly registered to the aligned volumes using a non-rigid registration based on a deformation sphere technique [10]. In the deformation sphere technique, smooth deformations are applied to voxels inside a sphere in such a way that the NMI is maximized. The location of the sphere is randomly chosen from the surfaces of ventricles, atria and epicardium, and it is varied during the iteration.

The nodes of the triangulated surface model of the reference subject, obtained from the manual segmentation, were considered as semi-landmarks. Semi-landmarks were used because only a few anatomical landmarks can be located from the heart in the MR images. Propagating the semi-landmarks, using the non-rigid transformations defined above, a set of corresponding semi-landmarks was achieved for each training set subject. The mean shape and its variance was then computed by applying Eqs. 2 and 3. In addition, the mean gray-scale short- and long-axis volumes were computed.

To reduce the bias of the mean shape towards the selected reference subject, and to give a better *a priori* estimate in the non-rigid registration, the preceding procedure was repeated by using the mean shape as a reference model.

In this work, two datasets were used: 1) a set of 2D contours from one long-axis image of each subject (Fig. 1a), and 2) a set of 3D triangulated surfaces from each subject (Fig. 1b). The number of data points was $P = 219$ in 2D and $P = 2086$ in

3D. The number of deformation modes to represent any arbitrary contour or surface is $D \cdot P = 2 \cdot 219 = 438$ or $3 \cdot 2086 = 6258$, respectively. Because the number of subjects is $N = 24$ (24 instead of 25 because of cross-validation) in this work, the maximum number of deformation modes is, however, 23 ($N - 1 < D \cdot P$) for the standard PCA.



(a) (b)

Fig. 1. *A set of a) 2D contours and b) 3D surfaces of atria, ventricles and epicardium from one subject.*

2.3 Techniques to increase the size of training set

Techniques to enlarge artificially the training set have been widely studied [11–15]. Next, the techniques tested in this work are summarized.

The values in parentheses in the text below are related to user-defined parameters and indicate the values that produced the lowest mean point-to-point error (Section 2.4). These values were used in Section 3.

Standard PCA. The procedure described in Section 2.1 was followed producing $N - 1$ deformation modes.

PCA & FEM. Cootes et al. [11] combined the standard PCA and finite element method (FEM) in shape modeling. FEMs take an instance of a shape and treat it as if it was made of flexible material. Modal analysis gives a set of linear deformations of the shape, such as bend, shear and pinch, equivalent to the modes of vibration of the original shape. These modes can be used in shape modeling:

$$\mathbf{x} = \mathbf{x}_i + \mathbf{\Omega}_i \mathbf{u}, \tag{6}$$

where $\mathbf{x}_i$ is an example of shape (as in Eq. 2), $\mathbf{\Omega}_i$ is a matrix consisting of eigenvectors computed for the stiffness matrix of the shape i, and $\mathbf{u}$ is a weight vector (as $\mathbf{b}$ in Eq. 1).

The deformation modes are the eigenvectors of the matrix

$$S = \mathbf{\Sigma} + \alpha(\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Omega}_i\mathbf{\Lambda}_i\mathbf{\Omega}_i^T), \tag{7}$$

where $\mathbf{\Sigma}$ is the covariance matrix from Eq. 3, α is a user defined weight ($\alpha = 0.5$), N is the number of chosen example shapes, $\mathbf{\Lambda}$ is an inverse matrix of the diagonal matrix consisting of eigenvalues of the stiffness matrix. The procedure produces $D \cdot P$ deformation modes.

Adaptive focus. In the standard PCA method, each point has an equal weight in computing the covariance matrix. Shen and Davatzikos [12, 13] proposed a method called adaptive focus where objects having a low spatial variance or objects with a high confidence were spatially scaled in order to increase the variance. Due to scaling, the deformations of the scaled object became more emphasized and better represented in the most important deformation modes (high eigenvalues because of high variance). In this work, each object was scaled separately in the co-ordinate system of the reference model (the scaling factor 2). The procedure produces $(O+1)N-1$ deformation modes where O is the number of objects ($O = 5$) and N is the size of the training set. The factor $O + 1$ is used instead of O because the original examples are also included in addition to scaled ones.

Non-rigid scaling. In this approach, the adaptive focus technique was extended to non-rigid but smooth deformations. The contours were scaled inside of a deformation sphere. The scaling factor, $s = s(x, y, z)$, of a point (x, y, z) is computed from

$$s(x,y,z) = \frac{e^{-2\frac{(x-c_x)^2+(y-c_y)^2+(z-c_z)^2}{r^2}} - e^{-2}}{1.0 - e^{-2}}S + 1, \tag{8}$$

where (c_x, c_y, c_z) and r are the location and the radius of the sphere ($r = 50$ mm), respectively, and S is the user specified scaling factor ($S = 1$). The sphere is randomly located to L locations ($L = 100$), and the original contour points are deformed at each location. The origin during the scaling is in the center of the sphere. The number of deformation modes produced is $(L + 1)N - 1$.

Non-rigid movement. Another strategy, very similar to the non-rigid scaling technique, was also tested. The displacement vector, $\mathbf{v}(x, y, z)$, for any point inside the sphere is computed from

$$\mathbf{v}(x,y,z) = \frac{e^{-2\frac{(x-c_x)^2+(y-c_y)^2+(z-c_z)^2}{r^2}} - e^{-2}}{1.0 - e^{-2}}\mathbf{V}, \tag{9}$$

where $\mathbf{V}$ is a random vector and other parameters as in Eq. 8. The length of the vector $\mathbf{V}$ was chosen from a uniform distribution ($[0\ \ 25]$ mm).

Fourier. The approach adopted in this work is closely related to the hierarchical method proposed in [14] where the data were divided into different frequency and spatial location bands using wavelets, and PCA was performed for each band separately. In this work, the data were decomposed only in frequency bands. The deviations of the training set examples from the mean were transformed into the frequency space using Fourier transformation and the data were band-pass filtered into B ($B = 18$) separate

frequency bands. New artificial training set examples were then generated by restoring each band separately into the shape space with inverse Fourier transformation. This procedure produces $(B+1)N-1$ deformation modes, when also original examples are included in the training set.

Noising. In this approach, Gaussian noise is added to each point which makes data less correlated. The displacement of a point was chosen from a uniform distribution ($[-2\ \ 2]$ mm) in each direction. From each training set example, L ($L = 100$) noisy contours were generated leading to $(L+1)N-1$ deformation modes. Alternatively, the data could be made more uncorrelated by replacing the covariance matrix $\mathbf{\Sigma}$ by $\mathbf{\Sigma} + \alpha\mathbf{I}$, where α is a weight factor and $\mathbf{I}$ a unit matrix.

2.4 Evaluation

As mentioned above, the segmentation accuracy does not depend only on the model properties but also details of the optimization method and image characteristics affect the result. In addition, normally no real gold standard exists for evaluating the accuracy of the segmentation. The automatic segmentation result is usually compared with the manual one which is commonly known to contain errors. Warfield *et al.* [16] recently proposed a solution to this problem. In our work, two methods were used in evaluation: 1) the model was fitted directly to a training set example using Eq. 5 (model-to-shape fit), and 2) the model was fitted iteratively to image data (model-to-image fit). The former method measures the ability of the model to represent an unseen object while the latter method includes all error sources in segmentation.

Two error measures are defined. Point-to-point (PP) error is computed as an average Euclidean distance between the corresponding points in $\mathbf{x}'$ and $\mathbf{x}$ (Eq. 4). The PP error can not normally be used in segmentation, because the point correspondences are not known. Therefore, point-to-curve/surface (PCS) error is defined: 1) search the shortest Euclidean distance from each point of $\mathbf{x}'$ to the contour or surface defined by $\mathbf{x}$, and 2) take an average of these distances. In other words, the PCS error omits the error in the tangential direction of the contour or surface, and produces lower values than the PP error. Both PP and PCS errors are used to measure the model-to-shape fit while only the PCS error is used with the model-to-image fit.

Cross-validation was used: each training set example was once regarded as a target, the shape model was built using the remaining training set, and the target was represented by the shape model.

3 Results

Model-to-shape fit. The 2D PP error (in [mm]) for different techniques is represented in Fig. 2. A non-parametric Wilcoxon Signed Ranks Test was used to detect statistically significant differences between the shape models when the number of deformation modes was 23. Statistical significance was considered to be obtained for p-value $p < 0.01$. The non-rigid movement technique produced the best result: as compared with the non-rigid scaling ($p < 0.01$) and with the other methods ($p < 0.0001$). This indicates that partly artificially generated modes perform better than the modes derived

from the original training set using the standard PCA. The curves show also that the PP error decreases very slowly after $100 - 150$ modes. In addition, the noising technique is clearly the worst approach. Fig. 3 shows the corresponding curve for the non-rigid movement technique as applied to 3D surfaces.

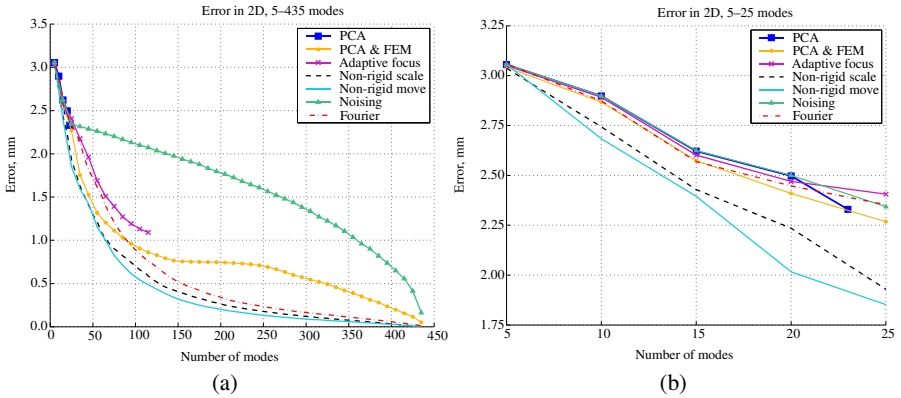
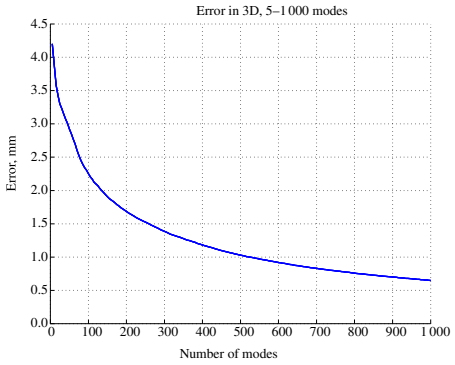


Fig. 2. *a) The 2D PP error in function of deformation modes for different techniques. b) The same curves only up to 25 modes.*

Model-to-image fit. Preliminary 3D segmentation results are also provided. The following method was used to optimize the weights of the deformation modes. For each data point ($P = 2086$), two gray scale profiles normal to the surface were generated: one from the mean short-axis and one from the mean long-axis data sets. The NMI was maximized between the profiles from the model and the profiles from same locations in the target data. All points from all profiles and from the both data sets were used in computing the NMI. In this work, the length of the profile was 21 points and the data sets were quantized to 32 gray values. Conjugate gradient method was used to optimize the weights of the deformation modes.

The segmentation results have been shown in Fig.3. No statistically significant differences were found between PCS values. However, when the NMI values were compared, the artificial modes were found superior compared with the standard PCA ($p < 0.00001$) and 100 or 250 artificial modes produced better results than 23 artificial modes ($p < 0.00001$). The improvement in NMI but not in the segmentation error indicates the existence of errors in our manual segmentation results used as a gold standard. The difference in the NMI values between 100 and 250 modes was not statistical significant. The segmentation result of one subject is visualized in Fig. 4.

The 3-D segmentation errors of cardiac structures, produced by automatic tools not based on the PCA-based approaches, have been recently reported to be around $2 - 3$ mm [9, 17]. The PCS error was about 55 % and 46 % of the PP error with our 2D and 3D data, respectively. This means that 23 deformation modes (3.6 mm PP error) have enough degrees of freedom to reach 1.7 mm PCS error. As our segmentation results



Method	PCS [mm]	NMI
PCA23	2.09 ± 0.43	1.086 ± 0.011
NRM23	2.04 ± 0.38	1.094 ± 0.012
NRM100	2.00 ± 0.34	1.102 ± 0.013
NRM250	2.04 ± 0.36	1.103 ± 0.014

Fig. 3. *The 3D PP error in function of deformation modes for the non-rigid movement technique. Table shows segmentation error and the corresponding NMI value for the standard PCA and non-rigid movement technique (NRM) with different number of deformation modes.*

indicated, the real segmentation error is higher: the PCS error was 2.09 mm for 23 modes defined by the non-rigid movement technique. For 250 modes, the PCS error was 2.04 mm while the minimum PCS error, due to limited degrees of freedom in the model, is only 0.7 mm (46 % of about 1.5 mm for 250 modes in Fig. 3).

4 Discussion

Different techniques to increase artificially the number of deformation modes were studied. The technique based on non-rigid movements produced the best results, also in statistical sense. The advantages of the method were that 1) the PP error decreased fastest as the size of the training set was increased, and 2) the PP error was lower than using the standard PCA with an equal number of modes. The latter point indicates that the standard PCA restricts the shape space too much if the training set is small, and introducing artificial variation improves the generality of the model. On the other hand, the artificially enlarged training set has also drawbacks. The model may become physically implausible and unrealistic segmentation results can be produced for low-quality and complex image data. In addition, statistical shape models can be used to detect different abnormalities from images if the training set is defined from healthy volunteers. With artificial training set this property is lost.

In order to avoid overfitting of the model, a common habit is to select only modes that explain, for example, 99 % of the variance in the training set. The more L-shaped the error curve is (Figs. 2 and 3), the easier the selection of the optimal number of modes is. The optimal number of modes is attained in the cross-section of vertical and horizontal parts of the L-curve because the error decrease per an added mode is small after that point [18]. In our cardiac cases, the cross-section point was approximately at 130 modes in 2D and at 250 modes in 3D as the non-rigid movement technique was used. In other words, the model does not improve considerably by adding more data to

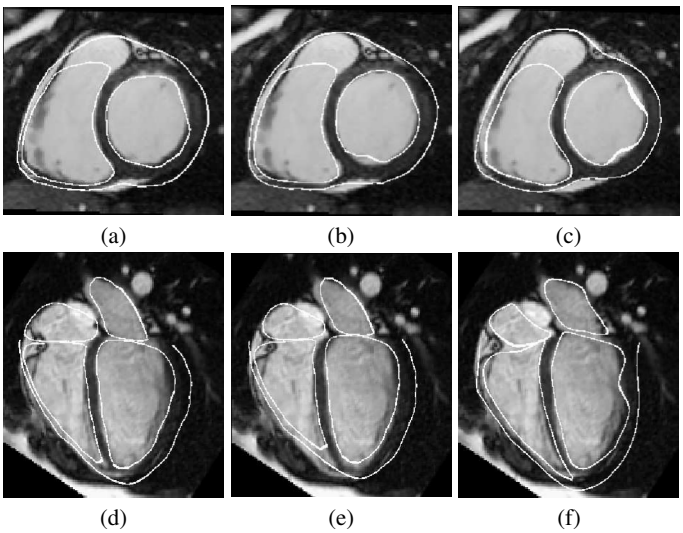


Fig. 4. *Segmentation result superimposed on MR slices of one subject. A short-axis view showing the result for a) the standard PCA (PCS error 1.95 mm), b) the non-rigid movement technique (PCS error 1.80 mm, 250 modes) and c) manual segmentation. The images d), e) and f) show the corresponding results in a long-axis view.*

the training set after these limits. Although these limits have been derived from a partly artificially generated training set, we believe that the shape of curves for real data would be approximately similar, and the limits computed from artificial data provide a rough estimate of the optimal size of the training set also for real data.

The results indicated that although the model with extended training set is capable to represent more accurately unseen objects, the segmentation accuracy does not improve equally. Two reasons explain the relatively small improvement in the segmentation accuracy: 1) a local maximum of NMI has been found because the conjugate gradient method is not a global optimization technique, and 2) the manual segmentation used as a gold standard contains errors. The relative contribution of these error sources should be studied in future.

Acknowledgements

This research was partly supported by the National Technology Agency, Finland.

References

1. Christensen, G.E., Miller, M.I., Vannier, M.W., Grenander, U.: Individualizeing neuroanatomical atlases using massively parallel computer. IEEE Computer **29** (1996) 32–38

2. Wang, Y., Staib, L.H.: Pysical model-based non-rigid registration incorporating statistical shape information. Med. Image Anal. **4** (2000) 7–20
3. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models - their training and application. Computer Vision and Image Understanding **61** (1995) 38–59
4. Davies, R., Twining, C., Cootes, T., Waterton, J., Taylor, C.: A minimum description length approach to statistical shape modeling. IEEE Trans. Med. Imag. **21** (2002) 525–537
5. Frangi, A., Rueckert, D., Schnabel, J., Niessen, W.: Automatic construction of multiple-object three-dimensional statistical shape models: Applications to cardiac modeling. IEEE Trans. Med. Imag. **21** (2002) 1151–1166
6. Kaus, M., Pekar, V., Lorenz, C., Truyen, R., Lobregt, S., Weese, J.: Automated 3-D PDM construction from segmented images using deformable models. IEEE Trans. Med. Imag. **22** (2003) 1005–1013
7. van Ent, D., de Munck, J., Kaas, A.: A fast method to derive realistic BEM models for E/MEG source reconstruction. IEEE Trans. Biomed. Eng. **48** (2001) 414–423
8. Rueckert, D., Frangi, A., Schnabel, J.: Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. IEEE Trans. Med. Imag. **22** (2003) 1014–1025
9. Lötjönen, J., Kivistö, S., Koikkalainen, J., Smutek, D., Lauerma, K.: Statistical shape model of atria, ventricles and epicardium from short- and long-axis MR images. Med. Image Anal. **8** (2004) 371–386
10. Lötjönen, J., Mäkelä, T.: Elastic matching using a deformation sphere. In Niessen, W.J., Viergever, M.A., eds.: Lecture Notes in Computer Science 2208: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2001, Springer (2001) 541–548
11. Cootes, T.F., Taylor, C.J.: Combining point distribution models with shape models based on finite element analysis. Image and Vision Computing **13** (1995) 403–409
12. Shen, D., Davatzikos, C.: An adaptive-focus statistical shape model for segmentation and shape modeling of 3-D brain structures. IEEE Trans. Med. Imag. **20** (2001) 257–270
13. Shen, D., Davatzikos, C.: An adaptive-focus deformable model using statistical and geometric information. IEEE Trans. Patt. Anal. Mach. Intell. **22** (2000) 906–913
14. Davatzikos, C., Tao, X., Shen, D.: Hierarchical active shape models using the wavelet transform. IEEE Trans. Med. Imag. **22** (2003) 414–423
15. de Bruijne, M., van Ginneken, B., Viergever, M., Niessen, W.: Adapting active shape models for 3d segmentation of tubular structures in medical images. In Taylor, C., Noble, A., eds.: Lecture Notes in Computer Science 2732: Information Processing in Medical Imaging, Springer (2003) 136–147
16. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. IEEE Trans. Med. Imag. **23** (2004) 903–921
17. Lorenzo-Valdés, M., Sanchez-Ortiz, G.I., Elkington, A.G., Mohiaddin, R.H., Rueckert, D.: Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. Med. Image Anal. **8** (2004) 255–265
18. Horn, J.L.: A rationale and test for the number of factors in factor analysis. Psychometrika **30** (1965) 179–186