

Artificially Enlarged Training Set in Image Segmentation

Tuomas Tölli¹, Juha Koikkalainen², Kirsi Lauerma³, and Jyrki Lötjönen²

¹Laboratory of Biomedical Engineering, Helsinki University of Technology, P.O.B. 2200, FIN-02015 HUT, Finland

`tuomas.tolli@tkk.fi`

²VTT Information Technology, P.O.B. 1206, FIN-33101 Tampere, Finland
{`juha.koikkalainen, jyrki.lotjonen`}@vtt.fi

³Helsinki Medical Imaging Center, University of Helsinki, P.O.B. 281, FIN-00029 HUS, Finland

Abstract. Due to small training sets, statistical shape models constrain often too much the deformation in medical image segmentation. Hence, an artificial enlargement of the training set has been proposed as a solution for the problem. In this paper, the error sources in the statistical shape model based segmentation were analyzed and the optimization processes were improved. The method was evaluated with 3D cardiac MR volume data. The enlargement method based on non-rigid movement produced good results – with 250 artificial modes, the average error for four-chamber model was 2.11 mm when evaluated using 25 subjects.

1 Introduction

Quantitative information, such as various volumetric measures, are needed to make more objective diagnosis. However, manual extraction of such information from images is both highly time-consuming and prone to segmentation errors due to subjective image interpretation. Therefore, accurate and reliable methods to automatically perform the segmentations are highly needed. Development of automatic segmentation methods has appeared to be challenging due to several reasons, such as high variability in the shape and appearance of various anatomical regions, low contrast between objects and noise in images.

Numerous model-based segmentation methods have been studied to solve the aforementioned difficulties. Especially active shape models (ASM) [1] has turned out to give promising results. In ASM, the mean model and its typical deformation modes, based on a training set, are defined using Principal Component Analysis (PCA). Several ASM-based approaches have been proposed for the 3D modeling and segmentation of medical images [2, 3, 4, 5]. Furthermore, the methods have lately been used to model the function of the heart [6].

Because the model construction process is time consuming and not enough data are always available, the number of examples in the training set remains often relatively low. This means that the degrees of freedom in the model are limited and the model can not represent accurately all cases in a population from

which the training set has been sampled. At least two techniques have been proposed to the problem: for example in [7, 8], the approach was to decrease the dimension of the model, whereas in [9, 10], the size of the training set was artificially increased.

In addition to testing different methods to enlarge the dataset artificially in [10], we studied the relation between the size of the training set and the ability of the model to represent an unseen object. We found that a heart model with 100 modes has enough degrees of freedom to reach about 1 mm segmentation accuracy for ventricles and atria. However, preliminary tests with real data showed that the segmentation accuracy did not improve as much as the modeling accuracy.

In this work, we identified three sources for the segmentation error: 1) the properties of the model, i.e., degrees of freedom in the model, 2) the properties of the deformation algorithm, such as goodness of criterion to be optimized and an optimization algorithm used, and 3) the errors in the golden standard to which a segmentation result is compared. Our objective is to study the contribution of these error sources to the final segmentation result. This was obtained 1) by proposing a method to eliminate the effects of the manual segmentation errors on the segmentation accuracy, 2) by improving and testing different optimization methods to deform the shape model and 3) by further analyzing and verifying the effects of the artificial enlargement of the training set. Image data used consisted of cardiac MR volumes.

Another contribution made in this work was the development of a novel segmentation algorithm for cardiac MR images. The method was based on artificially enlarged ASM. The weights of the deformation modes were optimized by fitting line profiles, defined for model points, to image data by simulated annealing.

2 Materials and Methods

Materials. The database consisted of both long- and short-axis MR volumes obtained from 25 healthy subjects. Atria, ventricles and epicardium were manually segmented from the volumes by an expert. The segmentations were made by fitting a triangulated surface model simultaneously into the short- and long-axis volumes. All manual segmentations were made using a specially designed software package [3].

Model Construction. Both the mean shape $\bar{\mathbf{s}}$ and the mean gray-scale short- and long-axis volumes were constructed according to the procedure proposed in [3]. Also, the point correspondence between the training set shapes \mathbf{s}_i were obtained, and used to model the shape variability in the training set using point distribution model (PDM) [1]. A covariance matrix was computed from the deviations of the training set shapes from the mean shape. The deformation modes, i.e. modes of shape variation, were defined by computing eigenvectors and -values of the covariance matrix.

In statistical shape models, new examples of the shape, $\mathbf{s} = [s_1, \dots, s_n]^T$, that are specific to the studied object, are generated using a linear combination

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{\Phi} \mathbf{b} , \quad (1)$$

where $\bar{\mathbf{s}} = [\bar{s}_1, \dots, \bar{s}_n]^T$ is a reference shape, typically a mean shape constructed from a training set, $\mathbf{\Phi} = [\phi_1, \dots, \phi_m]$ is a matrix consisting of the deformation modes ϕ_i , and $\mathbf{b} = [b_1, \dots, b_m]^T$ is a weight vector. The number of the non-zero eigenvectors is $\min(N - 1, n)$, where N is the number of the training set shapes and n the number of shape parameters, i.e. point coordinates in PDM.

To avoid the training set constricting too much the model deformation during the segmentation process, the size of the training set, and hence, the number of the deformation modes was artificially enlarged. In [10], we evaluated several methods and the best one called a non-rigid movement strategy was adopted here. The method moves non-rigidly surface points inside a deformation sphere. The sphere is randomly positioned to 50 locations on the object surfaces, and the original surface points are locally deformed at each position. The displacement vector for the movement is calculated from

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

where (c_x, c_y, c_z) and r are the location and the radius of the sphere ($r = 50$ mm), and \mathbf{V} a random vector. The length of the vector \mathbf{V} is chosen from a uniform distribution ([0 25] mm). Outside the sphere, the displacement vector is zero.

In ASM-based segmentation, the mean model is deformed according to Eq. 1 in such a way that the deformed model fits optimally to the target data. In other words, the objective is to find a combination of weights b_i optimizing a selected similarity function. In this work, the weights were constricted to satisfy $|b_i| \leq 5\sqrt{\lambda_i}$, in which λ_i is the eigenvalue corresponding to the i th eigenvector.

Optimization Criteria. Two optimization criteria were tested for the model deformation. Both criteria were based on the use of line profiles normal to the subject surface [11]. The first method utilized Normalized Mutual Information (NMI) [12]. For each surface point j , the grayscale information along the profile was sampled and put into the vector \mathbf{g}_j . The varying of the weight vector \mathbf{b} was made in such a way that the NMI between the profiles from the mean model and the profiles from the same locations in the target data was maximized.

In the second method, the normalized profiles \mathbf{g}_j were built for each surface point of each subject in the training set. Then, their mean $\bar{\mathbf{g}}_j$ and covariance \mathbf{S}_g were calculated. Here, it was assumed that the normalized samples were distributed as a multivariate Gaussian. Thereafter, the quality of fit between the target profiles and the profiles from the mean model was computed by minimizing the sum of the statistical (Mahalanobis) distances, $f(\mathbf{g})$, calculated over all the surface points as

$$f(\mathbf{g}) = \sum_{j=1}^n (\mathbf{g}_j - \bar{\mathbf{g}})^T \mathbf{S}_g^{-1} (\mathbf{g}_j - \bar{\mathbf{g}}) . \quad (3)$$

Optimization Algorithms. A commonly known problem with deformable models is that the deformation fails easily if the model is not close enough to the final result. In other words, the deformation stops to some local minimum/maximum of the optimization function. Therefore, we evaluated in this work both local and global optimization methods. Conjugate gradient and simplex optimization methods were used as local optimization methods. Simulated annealing (SA), a method simulating physical annealing process was used for global optimization. In SA, annealing schedule for the simulated temperature was calculated from

$$T = T_0 \left(1 - \frac{k}{K}\right)^\alpha , \quad (4)$$

where T_0 is the initial temperature, k the iteration number and K the number of total iterations ($\alpha = 4$). All the implementations pertaining to the optimization methods were based on the algorithms proposed in [13].

Modified Training Set. The segmentations were first done for original MR volumes with both normal PCA modes and artificially enlarged modes, using the mean model as an *a priori* model. However, the errors in the manual segmentations were discovered to have an influence on the process both during the deformation step (wrong shapes in the training set) and during the method validation step (inaccurate golden standard). Because no realistic cardiac MR phantom images were available, the following strategy was used to eliminate the errors in the golden standard.

A modified training set was built by exploiting the transformation fields obtained during the model construction phase. New *a priori* model was obtained by registering the database subject k with the mean model using the volumetric transformation $\mathbf{T}_k : x_k \mapsto \bar{x}$ (Fig. 1a). To better simulate the mean model, the new model was filtered with a Gaussian lowpass filter (Fig. 1b). Moreover, the target volumes were obtained by registering database subject k to all other database subjects l using the transformations $\mathbf{T}_{kl} : x_k \mapsto x_l$. This yielded subjects having the grayscale information from subject k and shapes from subject l . The segmentations were then applied to these volumes using the same statistical shape models than in normal MR volume segmentations. Thus, the manual segmentation errors were propagated both to the mean model and to the database subjects in a similar way, removing the effects on the measured segmentation accuracy.

Evaluation. First, the generality of the statistical shape model based segmentation was verified by applying the method to labeled volumes, where each object was presented by one grayscale value (Fig. 1c). Thereafter, the segmentation was evaluated both with 25 real and modified short- and long- axis MR volumes. In the segmentations, both the NMI- and Mahalanobis-based optimization criteria were used with different optimization algorithms. In addition to standard PCA,

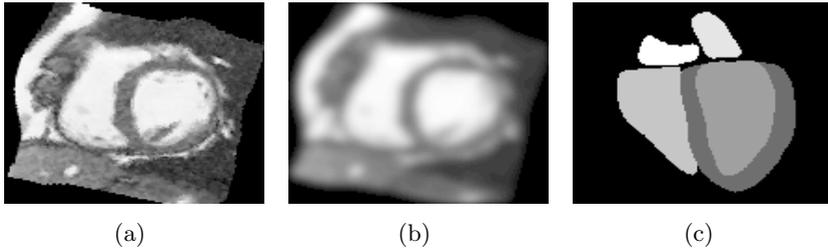


Fig. 1. An *a priori* model of the modified training set, i.e., one database subject transferred into the mean coordinate system a) before and b) after the Gaussian lowpass filtering process. c) An example of labeled volumes used to verify the generality of the statistical shape model based segmentation.

the segmentations were made using 24, 100 and 250 artificial modes, respectively. During all tests, leave-one-out cross-validation was used to avoid biasing.

Euclidean distances between the results obtained from the automatic and manual segmentation processes were used as error measures. The distances were averaged over all surface points of all subjects. Moreover, comparison between the absolute energy values were also made to better understand the phenomena.

3 Results

The segmentation results are summarized in Table 1. Moreover, the error distributions between different cardiac objects using simulated annealing and Mahalanobis distance based optimization are shown in Table 2. As an example, two randomly selected results are illustrated in Fig. 2, one in short- and one in long-axis view.

4 Discussion

As shown in Table 1, the performance of the method with labeled volumes follows the accuracy estimated in [10], i.e. the modeling accuracy in terms of the number of deformation modes. For example with 100 modes, approximately 1 mm error was produced since the deformation modes did not have more degrees of freedom to represent the full range of shape variations of the target object. Furthermore, the error increased to 1.36 mm when the modified volumes were used. This means that image characteristics, such as image noise and poorly visible interfaces between neighboring objects, increased error approximately 0.3-0.4 mm. This 1.4 mm can be considered as the real segmentation error of the algorithm. Finally, as the errors from the manual segmentations were also present, the error further increased by 0.5-1 mm, being 2.16 mm for real MR volumes using Mahalanobis distance based optimization.

For comparison, we segmented manually the left and right ventricle as well as epicardium from seven cases twice. The difference between these segmentations

Table 1. Top: Segmentation error and its standard deviation (mm) for simulated annealing. The standard PCA as well as the PCA with 24, 100 and 250 artificial modes were used (columns). The errors from top to bottom are for the lower limit of segmentation accuracy estimated in [10] (theoretical), labeled volumes (NMI based optimization), modified volumes (NMI) and real MR volumes using both NMI and Mahalanobis based optimization. **Bottom:** Comparison of optimization algorithms. Segmentation error and standard deviation (mm) was computed using PCA with 100 artificial modes.

Method	Standard PCA	Artificial 24	Artificial 100	Artificial 250
Theoretical	-	1.50	1.03	0.70
Labeled	1.45 ± 0.48	1.42 ± 0.40	1.02 ± 0.36	1.00 ± 0.39
Modified	1.86 ± 0.48	1.67 ± 0.39	1.36 ± 0.30	1.33 ± 0.31
NMI	2.47 ± 0.52	2.54 ± 0.57	2.40 ± 0.43	2.28 ± 0.41
Mahalanobis	2.23 ± 0.44	2.26 ± 0.50	2.16 ± 0.41	2.11 ± 0.45

Algorithm	Labeled	Modified	NMI	Mahalanobis
Conjugate gradient	1.17 ± 0.37	1.43 ± 0.31	2.45 ± 0.49	2.21 ± 0.41
Simplex	1.03 ± 0.35	1.38 ± 0.29	2.39 ± 0.37	2.17 ± 0.43
Simulated annealing	1.02 ± 0.36	1.36 ± 0.30	2.40 ± 0.43	2.16 ± 0.41

was about 1.4 mm. This value is in accordance with the estimates obtained in [3] (1.8 mm) and in [4] (1.2 mm). The accuracy of automatic segmentation (Table 2) was about 1.9 mm for the ventricles and epicardium. In other words, the accuracy should be improved by 0.5 mm in order to reach the human accuracy. However, according to our results increasing the number of deformation modes above 100-200 does not solve the problem but the optimization of the weights of the deformation modes becomes the problem. Although global optimization was used, the accuracy could not be significantly improved. Two reasons can explain the result: 1) the function to be optimized does not have the global

Table 2. Segmentation error for different cardiac objects using simulated annealing and Mahalanobis based optimization criterion. The error and its standard deviation (mm) are obtained when using standard PCA (top) and PCA with 250 artificial modes (bottom). Abbreviations LV, RV, LA, RA and EC stand for left ventricle, right ventricle, left atrium, right atrium and epicardium, respectively. As shown in Table 1, total errors were 2.23 mm for standard PCA and 2.11 mm for PCA with 250 artificial modes.

	LV	RV	LA	RA	EC
St PCA	1.89 ± 0.36	2.23 ± 0.64	2.67 ± 0.76	3.21 ± 1.67	1.96 ± 0.40
Artif 250	1.77 ± 0.36	2.13 ± 0.63	2.44 ± 0.85	3.17 ± 1.89	1.85 ± 0.51

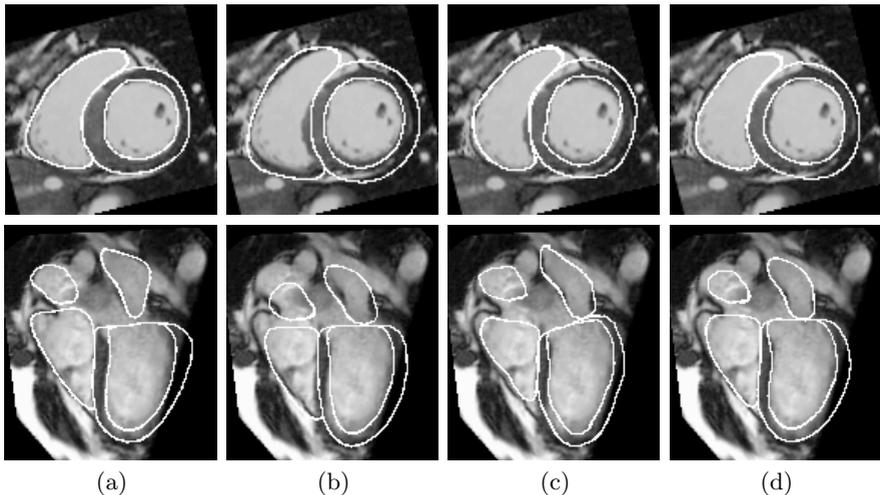


Fig. 2. Two randomly selected examples of the segmentations achieved by using Mahalanobis distance based criterion and simulated annealing. The slices with the short- (top) and long-axis (bottom) view have been chosen from different cases. a) Manual segmentations, b) initial conditions and c) automatic segmentations using standard PCA and d) PCA with 250 artificial modes. The errors for these particular cases were 1.65 and 1.61 (top) and 2.36 and 2.18 (bottom) for the standard PCA and PCA with 250 artificial modes, respectively.

maximum/minimum in the correct location due to image imperfections, and 2) the optimization was not run long enough to reach the global optimum. Despite the correct reason, increasing the computation time considerably is not feasible from clinical point of view. Therefore, more attention should be put on developing better optimization criteria. Furthermore, to ensure the artificial shapes produced by the non-rigid movement strategy (Eq. 2) would reflect true and physically plausible shape variation, it would be beneficial to constrain the deformations of the surface points using the local probability density functions generated from the database.

In terms of the energy optimization, simulated annealing outperformed simplex and conjugate gradient algorithms in all cases. However, since the difference in the segmentation errors was not statistically significant (Table 1), the choice between the optimization algorithm had no substantial role in the segmentation. The Mahalanobis distance performed slightly better than NMI when real MR volumes were used. This was probably due to small amount of information obtained from profile points.

For real MR volumes, the method developed resulted in relatively small final error, 2.11 mm by using 250 modes, simulated annealing and Mahalanobis-distance. The errors for the left and right ventricles were 1.79 mm and 2.14 mm, respectively. The errors are comparable with the results recently reported, and slightly better than reported for example in [3, 14].

References

1. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models - their training and application. *Computer Vision and Image Understanding* **61** (1995) 38–59
2. Mitchell, S.C., Bosch, J.G., Lelieveldt, B.P.F., van der Geest, R.J., Reiber, J.H.C., Sonka, M.: 3-D active appearance models: segmentation of cardiac MR and ultrasound images. *IEEE Trans. Med. Imaging* **21** (2002) 1167–1178
3. 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. *Medical image analysis* **8** (2004) 371–386
4. Assen, H., Danilouchkine, M.G., Frangi, A.F., Ordás, S., Westenberg, J.J.M., Reiber, J.H.C., Lelieveldt, B.P.F.: SPASM: A 3D-ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data. *Medical Image Analysis* **10** (2006) 286–303
5. van Ginneken, B., Frangi, A., Staal, J., ter Haar Romeny, B., Viergever, M.: Active shape model segmentation with optimal features. *IEEE Trans. Med. Imaging* **21** (2002) 924–933
6. Lorenzo-Valdéz, 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. *Medical Image Analysis* **8** (2004) 255–265
7. Davatzikos, C., Tao, C., Shen, D.: Hierarchical active shape models using the wavelet transform. *IEEE Trans. Med. Imaging* **22** (2003) 414–423
8. Zhao, Z., Aylward, S.R., Teoh, E.K.: A novel 3D partitioned active shape model for segmentation of brain MRI images. In: *Proc. MICCAI '05.* (2005) 221–228
9. de Bruijne, M., van Ginneken, B., Viergever, M., Niessen, W.: Adapting active shape models for 3D segmentation of tubular structures in medical images. In: *Proc. IPMI '03.* (2003) 136–147
10. Lötjönen, J., Antila, K., Lamminmäki, E., Koikkalainen, J., Lilja, M.: Artificial enlargement of a training set for statistical shape models: Application to cardiac images. In: *Proc. FIMH 2005.* (2005) 92–101
11. Cootes, T.: Statistical modes of appearance for computer vision. Tech. report, Univ. of Manchester (2004)
12. Studholme, C., Hill, D., D.J.Hawkes: An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition* **32** (1999) 71–86
13. Press, W., Teukolsky, S., Vetterling, W., Flannery, B.: *Numerical recipes in C: the art of scientific computing.* Second edn. Cambridge University Press (1992)
14. Kauss, M., von Berg, J., Weese, J., Niessen, W., Pekar, V.: Automated segmentation of the left ventricle in cardiac MRI. *Medical Image Analysis* **8** (2004) 245–254