

3D/4D Cardiac Segmentation Using Active Appearance Models, Non-rigid Registration, and the Insight Toolkit

Robert M. Lapp¹, Maria Lorenzo-Valdés², and Daniel Rueckert²

¹ Institute of Medical Physics, University of Erlangen-Nürnberg,
Henkestrasse 91, 91052 Erlangen, Germany

² Visual Information Processing Group, Department of Computing, Imperial College
London, 180 Queen's Gate, London SW7 2BZ, United Kingdom

Abstract. We describe the design of a statistical atlas-based 3D/4D cardiac segmentation system using a combination of active appearance models (AAM) and statistical deformation models with the Insight Toolkit as an underlying implementation framework. Since the original AAM approach was developed for 2D applications and makes use of manually set landmarks its extension to higher dimensional data sets cannot be easily achieved. We therefore apply the idea of statistical deformation models to AAMs and use a deformable registration step for establishing point-to-point correspondences. An evaluation of the implemented system was performed by segmenting the left ventricle cavity, myocardium and right ventricle of ten cardiac MRI and ten CT datasets. The comparison of automatic and manual segmentations showed encouraging results with a mean segmentation error of 2.2 ± 1.1 mm. We conclude that the combination of a non-rigid registration step with the statistical analysis concepts of the AAM is both feasible and useful and allows for its application to 3D and 4D data.

1 Introduction

A number of different segmentation approaches have been proposed that make use of statistical evaluations of the variations observed in one subject over time or across subjects. This allows to derive a probabilistic atlas of the possible variations which provides the means to produce *plausible* approximations of the objects stored. Recent applications of such statistical techniques to the domain of cardiac segmentation have been published [1,2,3,4] and also other new cardiac segmentation algorithms have been proposed [5].

1.1 Active Appearance Models

One promising statistical modeling approach are active appearance models (AAM), which were first proposed by Cootes *et al.* [6] in 1999 and have since received a lot of attention. In contrast to other techniques they utilize both shape and texture information of the object and lead to a unified description of possible *appearances* of the object.

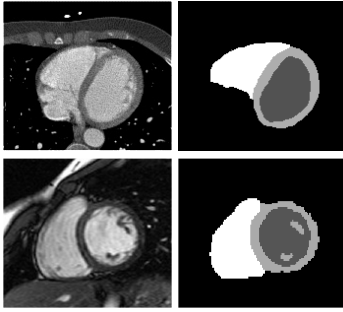


Fig. 1. CT and MRI input data and segmentations.

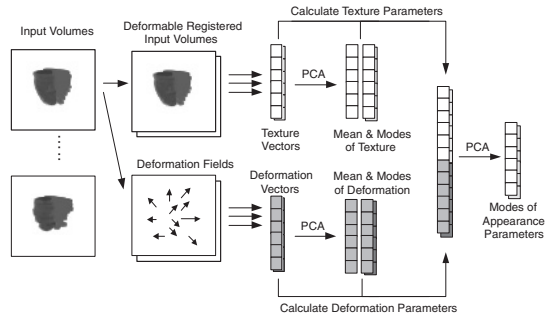


Fig. 2. Statistical atlas-building procedure.

Numerous applications have been published recently which apply the original concept with its manually identified landmarks to 2D segmentation problems and extensions to incorporate time have been proposed, as for example [7,8]. Also, a hybrid AAM and active shape model approach [9] was published recently for 2D, which demonstrates a fully automated segmentation of the left and right ventricle. However, these approaches are mainly applied to 2D data since they require a manual or semi-automatic localization of landmarks. Due to the lack of a sufficient number of uniquely identifiable points, establishing a dense map of point to point correspondences becomes significantly more difficult in higher dimensions.

Very few approaches for such higher dimensional data have been published and often depend on special geometrical models, which were customized for the specific application domain. For example, Mitchell *et al.* [1] show the application of AAM to 3D cardiac data. Their algorithm adapts a cylindrical model of the heart to the actual patient data thereby facilitating the identification of corresponding landmarks.

1.2 Statistical Deformation Models

Statistical deformation models (SDM) [10] aim to overcome this limitation and generalize the method of identifying corresponding points over various patients. They extend the concept of statistical shape models to allow their application in 3D and higher dimensions without manual landmarking. Instead, a deformable registration algorithm is used to register the reference object onto the training datasets. The resulting description of the dense mapping is then either used directly for the statistical analysis or is applied to transfer landmarks from one dataset onto the other. However, SDMs only consider the shape information of the object while the texture information is not utilized.

1.3 Contribution

We combined the concepts of active appearance motion models and statistical deformation models. To validate our ideas, we implemented a segmentation

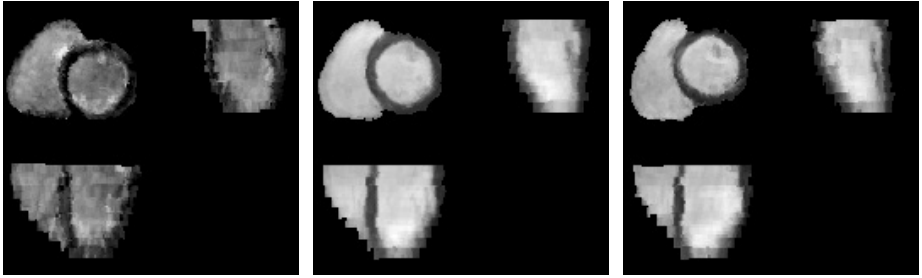


Fig. 3. Variation of the first mode of appearance parameters between $\pm 1.5\sigma$

system based on the Insight Segmentation and Registration Toolkit (ITK) [11] and applied it in 3D and 4D for the segmentation of the left ventricular (LV) myocardium, LV cavity and right ventricle (RV) of ten cardiac computed tomography (CT) and ten magnetic resonance imaging (MRI) datasets (cp. Fig. 1). For evaluation purposes we compared the volume overlap of the segmented voxels and calculated the mean distances between the automatic segmentation results and the manual segmentations which served as the gold standard.

2 Methods

2.1 Global Registration

Prior to the alignment of the data, the four dimensional training datasets and the corresponding manual segmentations are resampled to achieve the same 4D voxel and image size over all datasets. The training datasets are globally registered onto one arbitrary reference dataset using affine registration with an initial starting estimate generated by aligning the centers of mass. The registration is implemented within ITK's multi-resolution registration framework using a mutual information similarity measure and a linear interpolation scheme. In case of 4D datasets, these are split into their 3D time frames and registered separately. The mean transform over all time frames is subsequently calculated and applied to the 4D training datasets.

2.2 Local Registration

First, an average reference dataset for the local registration is created using an iterative procedure of deformable registering the datasets onto one (first arbitrary) dataset and subsequently calculating the mean appearance parameters. This new mean dataset is used as the new reference for the next iteration.

All training datasets are warped onto the reference subject using deformable registration. In particular, ITK's implementation of Thirion's *demons* algorithm [12] is applied which uses deformation fields as underlying representation, i.e. n -dimensional images with n -dimensional vectors as voxels which describe

the correspondence mapping of voxels between the training and reference images. These serve as the basis for the statistical analysis.

Two different approaches are available for deformable registration. In the first method, the datasets are split into their 3D time frames which are then registered separately. The resulting 3D deformation fields are assembled to a pseudo 4D deformation field with an additional zero element in the fourth dimension. This 4D field then actually describes a deformation where no influence between time frames exists. Alternatively, the non-rigid registration is directly performed on the 4D training datasets resulting in a genuine 4D deformation field.

The memory footprint of the resulting deformation fields is four times larger than the original image since the offset in every dimension has to be stored for every voxel. Therefore, a size reduction of the deformation field becomes necessary which is achieved by reducing the sample density by an arbitrary factor. Further memory saving measures include the introduction of automatic bounding boxes for deformation and texture regions.

2.3 Atlas Building Stage

As a first step, all training datasets and their segmentations are warped onto the standard shape using the result of the local registration, i.e. the deformation field. The texture information of every voxel from within the manually segmented standard shape is then concatenated to one texture vector \mathbf{t} . As shape information, the voxels of the deformation field within a bounding box around the LV and RV of the heart are concatenated to a deformation vector \mathbf{d} .

The subsequent statistical analysis follows the procedure as presented in the original AAM publication [6]. The principal component analysis (PCA) performs a dimensionality-reducing decomposition of the vectors into $\mathbf{t} = \bar{\mathbf{t}} + \mathbf{P}_t \mathbf{b}_t$ and $\mathbf{d} = \bar{\mathbf{d}} + \mathbf{P}_d \mathbf{b}_d$ with $\bar{\mathbf{t}}$ and $\bar{\mathbf{d}}$ being the mean texture and deformation vectors, \mathbf{P} being the orthogonal modes of variation and \mathbf{b} being the texture and deformation parameter vectors respectively. As a result, a mean model of shape and texture is created together with modes of variation, which capture the entire information about texture and shape of the training data. To incorporate both shape and greylevel information into one model both parameters \mathbf{b}_d and \mathbf{b}_t are concatenated and a further PCA is applied, i.e.

$$\begin{pmatrix} \mathbf{W}_d & \mathbf{b}_d \\ & \mathbf{b}_t \end{pmatrix} = \begin{pmatrix} \mathbf{Q}_d \\ \mathbf{Q}_t \end{pmatrix} \mathbf{c} = \mathbf{Q} \mathbf{c} \quad (1)$$

where \mathbf{W}_d is a diagonal matrix with weights allowing for different units in deformation and texture parameters, \mathbf{c} is the appearance parameters vector which controls both shape and greyscale of the model and \mathbf{Q} represents the resulting eigenvectors of the PCA. Because of the linear nature of the problem, \mathbf{Q} can be split into \mathbf{Q}_d and \mathbf{Q}_t as indicated above. This statistical atlas of the heart allows the creation of arbitrary plausible images of the heart within the variations of the training datasets. Fig. 2 gives an overview of the performed steps and Fig. 3 shows the variation of the first mode of variation of the appearance parameters between $\pm 1.5\sigma$ of one example cardiac MRI atlas.

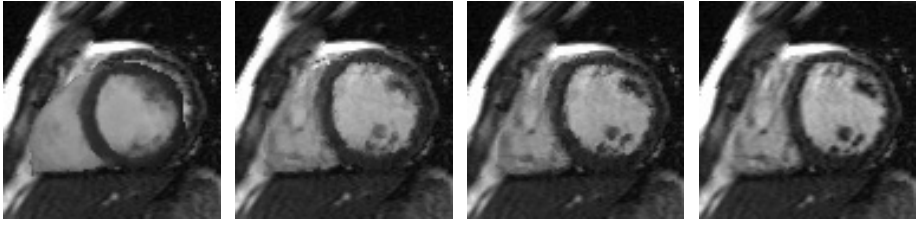


Fig. 4. Current hypothesis image within the enclosing target image while optimizing the appearance parameters: initial approximation, after few iterations, final approximation, target image (from left to right).

2.4 Image Segmentation

The statistical atlas is subsequently used for the actual image segmentation procedure. Given the results from before, any plausible image can be described by one appearance vector \mathbf{c} using the following equations

$$\mathbf{d} = \bar{\mathbf{d}} + \mathbf{P}_d \mathbf{W}_d^{-1} \mathbf{Q}_d \mathbf{c} \quad (2)$$

$$\mathbf{t} = \bar{\mathbf{t}} + \mathbf{P}_t \mathbf{Q}_t \mathbf{c} . \quad (3)$$

To synthesize new images a shape-free texture image is created from the resulting vector \mathbf{t} and warped by the deformation field \mathbf{d} . An iterative optimization process is now applied to adapt the appearance parameters to the actual image. A root-mean-square measure is calculated to determine the correspondence between the target image and the approximated image. The gradient information, required by the gradient descent optimization method, is calculated by forward differencing. Fig. 4 shows exemplary an iterative approximation of the shape and texture during the appearance optimization process. Finally, to obtain the segmentation result, the deformation field which is defined by the optimized appearance parameter set is used to deform the standard segmentation. The result is the approximated segmentation for the given target dataset.

3 Results

The performance of the system was evaluated using 10 MRI and 10 CT datasets. The cardiac MRI short-axis datasets had an original pixel size of between $1.37 \times 1.37 \times 10 \text{ mm}^3$ and $1.48 \times 1.48 \times 10 \text{ mm}^3$ and between 10 and 19 time frames each. The MRI cardiac short-axis datasets were acquired at the Royal Brompton Hospital, London, UK in a study with 10 healthy volunteers. A 1.5T Magnetom Sonata scanner (Siemens Medical Solutions, Erlangen, Germany) with a TrueFisp sequence was used with a total acquisition time of 15 minutes. For eight datasets manual segmentations and for two datasets semi-automatic segmentations were available.

The CT datasets were acquired during normal clinical practice using a Sensation 16 scanner (Siemens Medical Solutions, Forchheim, Germany) with a rotation time of 0.42 seconds and 12 slices acquired simultaneously. They were

kindly provided by the Institute of Diagnostic Radiology and the Department of Internal Medicine II, University of Erlangen-Nürnberg, Germany. The rawdata was reconstructed using a dedicated cardiac CT image reconstruction software (VAMP GmbH, Möhrendorf, Germany) using retrospective gating with 5 reconstructed cardiac phases and an initial isotropic voxel size of $(2\text{ mm})^3$. One CT dataset was segmented manually in 3D to provide an initial reference segmentation. To be able to calculate quantitative measures, also single slices of all other CT datasets were segmented manually.

3.1 Quantitative and Qualitative Performance

The performance of the system was evaluated qualitatively by visual inspection and quantitatively using two different measures. The distance between the manual reference segmentation and the automatic segmentation was measured in 3D and used to calculate maximum, mean and standard deviation of the absolute distance of surface points of the segmented volume to the closest point of the reference segmentation.

Also voxel volume percentages of correctly assigned voxels were calculated and used for evaluation. The measure *volume overlap* Δ was defined as

$$\Delta = \min \left(\frac{V_{M \cap A}}{V_M}, \frac{V_{M \cap A}}{V_A} \right) \quad (4)$$

with V_M being the volumes of the true (manual) segmentation, V_A the volume of the automatic segmentation and $V_{M \cap A}$ the volume of the correctly labeled regions, i.e. where the automatically computed segmentation values match the manual reference segmentation. While this volume measure is easily calculated and gives an indication of the quality of volumetric accuracy, the distance measure gives a more intuitively understandable result of the quality of the segmentation.

Figs. 5 and 6 show the results of the quantitative evaluation after global registration and segmentation using leave *one* out and leave *none* out schemes using all eigenmodes. The correlation for the segmented volumes for LV cavity, myocardium and RV in the MRI datasets is $r = 0.95$, $r = 0.83$ and $r = 0.94$, respectively. The mean absolute distance measure shows a segmentation error of 2.2 ± 1.1 mm for the leave-none-out scheme, which increases to 4.0 ± 1.6 mm for the leave-one-out test. For the CT segmentation, the volume overlap was calculated for one slice per volume only since no 3D segmentations were available. The evaluations show only a slight increase in segmentation accuracy when compared to the initial segmentation estimate which in our case originated mainly from poor deformable registration results.

The automatic segmentation of a typical $256 \times 256 \times 128$ volume with 8 time frames took approximately 20 minutes on a standard PC with 3 GHz and 2 GB main memory.

MAD MRI [mm]	Mean	Std Dev	Max
Initialization	4.6	2.3	56.1
AAM (one out)	4.0	1.6	29.4
AAM (none out)	2.2	1.1	25.4

Overlap MRI	LV cav	LV myo	RV
Initialization	0.66	0.42	0.71
AAM (one out)	0.72	0.49	0.73
AAM (none out)	0.86	0.68	0.86

Overlap CT	LV cav	LV myo	RV
Initialization	0.82	0.57	0.72
AAM (one out)	0.77	0.59	0.72
AAM (none out)	0.85	0.65	0.73

Fig. 5. Achieved segmentation accuracy in terms of mean absolute distance (MAD) and volume overlap.

4 Discussion and Conclusion

We developed and implemented a statistical atlas-based 3D/4D cardiac segmentation system based on the ITK and incorporating both texture and shape information. The problem of landmarking was solved by using the idea of statistical deformation models, i.e. using deformable registration to obtain point-correspondence information and taking the deformation information as shape description.

The implemented system proved to achieve satisfactory segmentation results with mean errors of about 2.2 ± 1.1 mm for the cardiac MRI images. The accuracy is comparable to [2] and slightly worse than the results presented by Kaus *et al.* [3]. Nevertheless, the full utilization of the time dimension did not show the expected improvement of the segmentation results, which was probably due to the limited temporal resolution of the datasets. Also, when working with high-resolution data, memory problems originated from the large texture vectors. This matter leaves room for improvement and will be investigated further.

The extensive use of the ITK leaves a positive impression and can especially be recommended for new researchers working in the field of image analysis. A migration to ITK may also be worthwhile because of its good architecture, multi-platform capabilities and the prospect of a broad developer basis. However, at many points a good command of C++ is required, especially for optimizing the relatively large memory footprint and the execution speed of the algorithms. One aspect that could be criticized is the software documentation. Nevertheless, there has been much improvement over the last months and a lot of effort has been put into the excellent software guide [11].

Interesting aspects of future work include the evaluation of other machine learning algorithms as e.g. the recent *locally linear embedding* approach [13] or *support vector machines* [14]. Desired properties include a better preservation

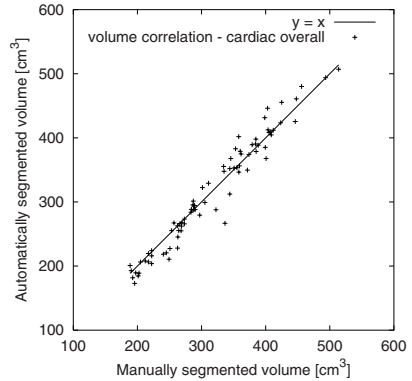


Fig. 6. Correlation of segmented volumes of MRI datasets using AAM segmentation and manual segmentation.

of neighborhood information and improved results for small numbers of training datasets. Also, CT and MRI datasets could be combined to a multi-modality atlas offering an increased number of available training datasets and cross-training, i.e. the refinement of the atlas for one modality by another imaging modality. Thereby, specific advantages of each modality, e.g. the high isotropic resolution of CT images, could be incorporated into a common atlas.

Acknowledgements. This work benefited from the use of the ITK, an open source software developed as an initiative of the U.S. National Library of Medicine. R.M. Lapp is funded by a grant from the DFG, Germany. M. Lorenzo-Valdés is funded by a grant from CONACyT, México.

References

1. Mitchell, S., Bosch, J., Lelieveldt, B., Geest, R., Reiber, J., Sonka, M.: 3-D active appearance models: Segmentation of cardiac MR and ultrasound images. *IEEE Trans. Med. Imag.* **21** (2002) 1167–1178
2. Lorenzo-Valdés, M., Sanchez-Ortiz, G., Mohiaddin, R., Rueckert, D.: Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. In: *MICCAI 2003*, Springer (2003) 440–450
3. Kaus, M., Berg, J., Niessen, W., Pekar, V.: Automated segmentation of the left ventricle in cardiac MRI. In: *MICCAI 2003*, Springer (2003) 432–439
4. Stegmann, M.B., Ersbøll, B., Larsen, R.: FAME - a flexible appearance modeling environment. *IEEE Trans. Med. Imag.* **22** (2003) 1319–1331
5. Noble, N., Hill, D., Breeuwer, M., Schnabel, J., Hawkes, D., Gerritsen, F., Razavi, R.: Myocardial delineation via registration in a polar coordinate system. In: *MICCAI 2002*, Springer (2002) 651–658
6. Cootes, T., Beeston, C., Edwards, G., Taylor, C.: A unified framework for atlas matching using active appearance models. In: *International Conference on Information Processing in Medical Imaging*. LNCS, Springer (1999) 322–333
7. Sonka, M., Lelieveldt, B., Mitchell, S., Bosch, J., Geest, R., Reiber, J.: Active appearance motion model segmentation. In: *Second International Workshop on Digital and Computational Video*, IEEE (2001)
8. Bosch, J., Mitchell, S., Lelieveldt, B., Nijland, F., Kamp, O., Sonka, M., Reiber, J.: Automatic segmentation of echocardiographic sequences by active appearance motion models. *IEEE Trans. Med. Imag.* **21** (2002)
9. Mitchell, S., Lelieveldt, B., Geest, R., Bosch, H., Reiber, J., Sonka, M.: Multistage hybrid active appearance model matching: Segmentation of left and right ventricles in cardiac MR images. *IEEE Trans. Med. Imag.* **20** (2001)
10. Rueckert, D., Frangi, A., Schnabel, J.: Automatic construction of 3D statistical deformation models of the brain using non-rigid registration. *IEEE Trans. Med. Imag.* **22** (2003) 1014–1025
11. Ibanez, Schroeder, Ng, Cates: *The ITK Software Guide*. Insight Consortium. (2003) <http://www.itk.org>.
12. Thirion, J.P.: Non-rigid matching using demons. In: *Computer Vision and Pattern Recognition*. (1996)
13. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. *Science* **290** (2000)
14. Golland, P., Grimson, W., Shenton, M., Kikinis, R.: Small sample size learning for shape analysis of anatomical structures. In: *MICCAI 2000*, Springer (2000) 72–82