AUTOMATIC ACTIVE APPEARANCE MODEL SEGMENTATION
OF 3D ECHOCARDIOGRAMS

K. Y. Esther Leung1, Marijn van Stralen1,2, Gerard van Burken1, Nico de Jong1, Johan G. Bosch1

1Thoraxcenter Biomedical Engineering, Erasmus MC Rotterdam, the Netherlands
2Div. Image Processing, Radiology, Leiden University Medical Center, the Netherlands

ABSTRACT

A fully automated segmentation for 3D echocardiography (3DE) using 3D Active Appearance Models (AAM) was developed and evaluated on 99 patients. The method used ultrasound specific grey value normalization and two matching algorithms were tested. To our knowledge this is the first report on a fully operational 3D AAM employed in 3DE on a large scale. The 3D AAM detected the endocardial contours accurately, even in the presence of large variations in left ventricular appearance and shape. Matching was successful in 91% of patients and resulted in a median point-to-surface error of 2.69 mm (av±sd: 2.91±1.03mm). Results indicate that fully automated AAM analysis of 3DE is practically feasible in datasets of mixed origin and quality.

Index Terms— Automatic segmentation, active appearance models, 3D echocardiography, left ventricular volume, ultrasound

1. INTRODUCTION

3D echocardiography (3DE) offers new possibilities in the analysis of cardiac function. It allows accurate assessment of left ventricular (LV) functional parameters such as LV volume, but common analysis methods for 3DE are still semiautomatic [1,2,3] and thus labor intensive and subjective. Fully automated methods (e.g. [4]) save valuable time in the analysis and eliminate observer variability.

We aim at fully automatic segmentation of the left ventricle in 3DE using 3D active appearance models (AAMs) [5]. AAM segmentation has proven to be successful in face recognition [5] and medical image segmentation, e.g. for 2D ultrasound sequences and 3D MRI [6]. However, for 3DE so far only preliminary results have been reported. In [6] a preliminary 3D AAM implementation was applied to 2D+T ultrasound. Recently, we reported on preliminary development [7] and on partial testing [8] of a 3D AAM for 3DE. No other reports have appeared yet, presumably due to the practical complexities in generating 3D AAMs from large datasets, and the complex regression training required for efficient 3D detection.

2. METHODS

We propose a fully automatic 3D AAM segmentation approach for the LV in 3DE. The 3D AAM implementation was developed in our lab, and described in [7,8] in detail. Two AAM matching algorithms were evaluated: the regular matching [5], and the jacobian tuning (JT) algorithm [9].

2.1. Active Appearance Models

AAMs represent the shape and the texture of an object such as the LV as a mean appearance with its eigenvariations, by applying principal component analysis (PCA) on example training data annotated by experts. We describe the training samples i by their shape \( s_i = (x_{i0}, y_{i0}, z_{i0}, ..., x_{iS}, y_{iS}, z_{iS}) \) containing \( S \) corresponding 3D surface points \( (x, y, z) \) and their texture \( t_i = (g_{i0}, ..., g_{iT}) \) containing \( T \) corresponding image samples \( g \). By applying PCA models are created, in which shape and texture are decomposed into their mean \(<s>\) and \(<t>\) and eigenvector matrix \( \Phi_s \) and \( \Phi_t \) [5]. Any shape or texture can be approximated in this model space with the parameter vector \( b_s \) and \( b_t \):

\[
s = \langle s \rangle + \Phi_s b_s \quad \text{and} \quad t = \langle t \rangle + \Phi_t b_t. \tag{1}
\]

Shape and texture are combined into appearance by concatenating \( b_s \) and \( b_t \) into the vector \( b \): \( b^T = (b_s^T \mid b_t^T) \) and applying a third PCA to find the appearance vector \( c \):

\[
b = \Phi_b c. \tag{2}
\]

We model the pose with a vector \( m \) using 7 parameters (1 scaling, 3 rotation, 3 translation). We combine appearance and pose in a total parameter vector \( p^T = (c^T \mid m^T) \) (\( T \): transposition).

AAMs are matched to unseen data by iteratively updating parameter vector \( p \). First, the model is initialized with average parameters \( p \) and placed in the image. Parameter updates are determined by evaluating the difference between the residual vector \( r(p) = t_p - t_m \), where \( t_m \) is the modeled texture as a function of the current parameters \( p \) (equation (2)) and \( t_p \) is the texture, sampled from the current location of the model’s shape in the image (and therefore also depends on \( p \)). The update process which minimizes \( r^T r \), is driven by the update matrix \( R \), which is the pseudo-inverse of the jacobian \( J = \frac{\partial r}{\partial p} \). \( J \) is estimated during the training stage by applying perturbations in the parameters \( p \) and finding the typical changes in \( r(p) \) over the whole training set by regression.
In the regular matching, \( R \) is fixed. In the JT approach, \( R \) is adapted to the image during the matching process. At each iteration, all previous parameter estimates and residuals are used to linearly update \( J \). The updated \( J \) is then used to update the parameters \( p \). For mathematical details, see [9].

2.2. Realization

The described methods have been developed in C++, using the ITK 3.8.0, VTK 5.0.4 and QT 4.4.2 toolkits. All AAM processing code was developed in-house.

Instead of using the standard (Cartesian) image coordinates, we sample our LV shapes and textures following a cylindrical/spherical anatomical coordinate system of the LV, oriented around the long axis (LAX) [7]. In this manner, the endocardial surface is represented with a regular sampling over equidistant angles \( a \) and long-axis heights \( h \), containing \( S = 901 \) 3D vertices (2703 coordinate values), see Fig. 1A.

We also sample the texture radially on lines through the surface points, from the LAX up to twice the radius of the surface. In this way, the myocardium and lumen, but also part of the right ventricle and a small region outside of the heart are modeled, enlarging the lock-in region of the AAM. Texture is represented by \( T = 22500 \) intensity samples.

For applying a PCA on the grey values of texture samples, a Gaussian distribution is assumed. It is known that the grey value distribution in ultrasound images is highly non-Gaussian. Therefore we apply a non-linear grey value normalization, as introduced in [10]. This normalization converts the grey values via a look-up table to an approximately Gaussian distribution.

3. EVALUATION SETUP

In this evaluation study we used end-diastolic (ED) images from 99 patient datasets, provided by TomTec Imaging Systems GmbH. The images (apical 3D views of the LV), were acquired with GE-Vingmed or Philips imaging systems in different hospitals and showed a variety of pathology, image appearance and quality. They were semiautomatically analyzed using TomTec’s 4D LV-Analysis. The tracing convention included the papillary muscles into the LV lumen, and excluded the trabeculations on the lateral part of the LV wall. This common approach in ultrasound tracing allows good evaluation of lateral wall contractility. It should be noted that such contours often do not correspond to the brightest edge in the images.

Initially, a model was built with all 99 patient sets (leave-0-out), to study the appearance variations and for perturbation experiments. The average appearance is shown in Fig. 1B, the first two eigenvariations are shown in Fig. 2. The first variation shows a transition from a long, narrow LV to a short, wide LV. Variations in the presence of the near-field artifact near the apex can be seen in the second variation. We successfully tested the generalization of the model and the basic matching convergence. For this leave-0-out model, perturbations up to 4 SD led to near-perfect convergence.

The ‘real-life’ segmentation evaluation was performed in a leave-8-out manner on the 99 datasets. The set was split in 13 groups of \( \leq 8 \) patients. Each group was used for testing with a model generated from the remaining 91 patients. Results were composed from all groups.
4. RESULTS

The leave-8-out models could generally represent the non-modeled objects adequately, such that any perturbations within 3SD would lead to an average point-to-point error (P2P) of about 2.7 mm (Fig. 3). Removing any lower eigenvariations from the leave-8-out models degraded the matching performance, suggesting that generalization is not yet optimal with 91 patients. In all further experiments, we used model variation coverage of 100%.

In the matching experiments, the model was initialized with average parameters at the image center and at a 25% higher vertical position. The latter was necessary to prevent mismatches since several images also contained the left atrium. The best match, determined by the final residual $r^T r$, is then taken.

Fig. 3. Perturbation experiments of the leave-8-out model on 99 patient sets (100% variation coverage). Results from regular matching and jacobian tuning were not significantly different.

This automated matching proved successful in the vast majority of cases. The detected contours follow the ground truth quite well. An example is given in Fig. 4. Some mismatches still occurred, similar to other AAM applications. Most of these were very abnormal or poorly centered ventricles. We classified any detection with average point-to-surface (P2S) error $>6$mm as a mismatch and found that 91% (90 of 99 patients) was detected successfully. The detection errors and volume results over the successful matches are summarized in Table 1 and Fig. 5. An average P2S error of $2.91 \pm 1.03$mm was found.

5. DISCUSSION

The automated detection performed well, the match was successful in 91% of the cases. A good correspondence to ground truth contours was found, both in terms of point-to-point errors and of volumes. This is especially encouraging, given the presence of large variations in image appearance and the mixed origin of the data. Note that these results were generated fully automatically on unseen data and can be further improved by more precise initialization and subsequent segmentation refinement. As we found in previous studies [10], the globally plausible AAM solution may easily be improved by dedicated local border refinement.

Table 1. Errors of detected LV shape and volume of the 90 successfully matched patients.

<table>
<thead>
<tr>
<th>POINT-TO-SURFACE (mm)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>average ± sd</td>
<td>2.91 ± 1.03</td>
<td>median</td>
<td>2.65</td>
<td></td>
</tr>
<tr>
<td>25% and 75% percentiles</td>
<td>[2.27 3.35]</td>
<td>25% and 75% percentiles</td>
<td>[5.43 8.96]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>POINT-TO-POINT (mm)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>average ± sd</td>
<td>7.49 ± 2.65</td>
<td>median</td>
<td>2.69</td>
<td></td>
</tr>
<tr>
<td>25% and 75% percentiles</td>
<td>[5.43 8.96]</td>
<td>25% and 75% percentiles</td>
<td>[5.43 8.96]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VOLUME (ml)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>regression</td>
<td>$y = -2.87 + 1.01x$</td>
<td>R</td>
<td>0.946</td>
<td></td>
</tr>
<tr>
<td>average ± sd</td>
<td>-1.47 ± 20.08ml</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.1. Regular vs. Jacobian Tuning Matching

In contrast to earlier findings [8], we did not find better model convergence by jacobian tuning for this data. We assume this is caused by suboptimal model generalization and a highly irregular solution space. Since the jacobian matrix is updated in JT during each matching step, regardless of whether the update improved the match, it can jump over local minima. However, this also means that it may diverge to an implausible solution, which seemed to be the case in some of our data sets. In this study, the regular algorithm was more stable and sufficiently accurate.

5.2. Appearance modeling and ground truth delineation

Fig. 4. Results of automated AAM matching. Green: ground truth contours, red: detected contours. Orthogonal cross sections through 3D dataset.
From visual inspection and perturbation experiments (Fig. 3), it appears that the models are adequate representations of the LV shape and texture. The lower bound of 2.7mm P2P can be attributed to anatomical variations in the patient data, but also to the variabilities in contour delineation and the definition of point correspondence. It appears that the number of datasets is not quite enough to cover the large amount of variations here; further study is needed to determine the optimal number, given the variations in imaging systems, hospital protocols, and image quality. Interobserver variability of the semiautomated analyses in this set is unknown, but in previous studies [11], we showed that variabilities in the annotation of anatomical landmarks in this type of data can easily amount to 4-8mm P2P, even with standardized annotation protocols. In comparison with other 3DE segmentation studies, our method performs very well also: e.g. Hansegård [3] reports ED volume errors of 2.1±10ml. The fully automated approach proposed by Orderud et al. [4] reported P2S of 2.2±0.4mm and ED volume errors of 3.6±11ml. Ground truth, dataset and algorithm issues make a direct comparison impossible, but it is clear that our method compares favorably. Since the 3D AAM models expert tracings and can incorporate typical artifacts, it can perform better in clinical quality 3DE than edge-based segmentation approaches. Local refinements and direct comparison to other approaches are subjects for further study.

6. CONCLUSIONS

We showed that 3D AAMs can indeed provide very promising segmentation of the LV in 3DE at ED. The algorithm gives a plausible, expert-like contour, even under varying acquisition settings. The method is fully automated.

7. ACKNOWLEDGMENTS

We thank Dr. M. Schreckenberg, C. Galuschky, and A. Nagy of TomTec Imaging Systems GmbH for providing the image and shape data in this study. This research is supported by the Dutch Technology Foundation STW (grant 06666), applied science division of NWO and the Technology Program of the Ministry of Economic Affairs.

7. REFERENCES