DATABASE GUIDED DETECTION OF ANATOMICAL LANDMARK POINTS IN 3D IMAGES OF THE HEART

Thomas Karavides¹,², K. Y. Esther Leung², Pavel Paclik³, Emile A. Hendriks¹, Johan G. Bosch²

¹Information and Communication Theory Group, Delft University of Technology, the Netherlands; ²Thoraxcenter Biomedical Engineering, Erasmus MC Rotterdam, the Netherlands; ³PR Sys Design, Delft, the Netherlands

ABSTRACT

Automated landmark detection may facilitate the examination and automatic analysis of three-dimensional (3D) echocardiograms. By detecting 3D anatomical landmark points, the standard anatomical views can be extracted automatically, for better standardized visualization. Furthermore, the landmarks can serve as an initialization for other analysis methods, such as segmentation. The described algorithm applies landmark detection in perpendicular planes of the 3D dataset. It exploits a database of expert annotated images, using an extensive set of Haar features for classification. The detection is performed using two cascades of Adaboost classifiers in a coarse to fine scheme. The method can detect landmarks accurately in the four-chamber (apex: 7.9±7.1mm, mitral valve center: 4.8±2.3mm) and two-chamber (apex: 7.1±6.7mm, mitral valve center: 5.2±2.8mm) views.

Index Terms—Classification, pattern recognition, landmark detection, ultrasound

1. INTRODUCTION

Real-time 3D echocardiography offers new possibilities in the analysis of cardiac function. The interpretation of the images, however, is critically dependent on the standardization of views. Variability in choice of anatomical views results in high interobserver and interinstitution variability in diagnosis. Additionally, quantitative analysis methods, like segmentation, often need a good initialization. A method for automatic detection of left ventricular landmarks (apex, mitral valve) would be of great importance.

Several methods for automated landmark detection in echocardiography have been proposed. Van Stralen et al. [1] describe detection of the left ventricular (LV) long axis and mitral valve plane in 3D ultrasound images. In 2D slices perpendicular to the acquisition axis, a circular Hough transform detects candidates for the endocardial border center. Dynamic programming finds a path through all slices that approximates the LV long axis. Further detection steps provide the position of the apex and mitral valve ring.

The Fourier Mellin Transform (FMT) has also been proposed for LV localization [2]. An appearance template of the LV is built, and both the template and the search image are transformed into a rotation and scale invariant representation using the FMT. This way, the estimation of rotation, scale and position converts into a cascade of linear shift detections, enabling a fast estimate of the object pose.

Recently, classification approaches have been proposed. Objects are represented using large sets of image features and detected by a classifier trained from labeled examples. Georgescu et al. [3] introduce a classification method to detect structures of interest in ultrasound images. A robust classifier is constructed using a boosted cascade of weak classifiers (a feature from a large set of Haar features). A similar approach is used in [4] for the detection and measurement of fetal structures in ultrasound images.

Lu et al. [5] presented a classification approach for the detection of standard view planes, similar to the one proposed in the current study. A cascade of three classifiers is trained based on boosting techniques. The classifiers use Haar wavelet-like feature types and steerable features [6].

In our study, a classification based method was developed to detect the standard anatomical landmark points (apex, mitral valve points) in two-chamber (2C) and four-chamber (4C) view planes in a fully automated way. The method uses smaller databases than those that have been used in similar approaches [5]; to make the classifier more robust, artificial variations are used. Our purpose is to use the proposed technique within an existing manual iterative technique for 3D landmark finding that uses 2D planes, so we can use 2D features to find the 3D points. Boosting techniques are used to train the classifier. Similar to [7,8] the method is applied to single 3D images at end-diastole (ED).

Fig. 1. Manual iterative 3D landmark detection approach. a) apex and mitral valve markers are placed in an approximate 4C plane, LAX is calculated. b) LAX is adjusted in perpendicular 2C approximation plane, c) perpendicular 4C plane is used to adjust LAX again, d) repeat b) and c) until true LAX found.
2. MATERIALS AND METHODS

2.1. Manual scheme

Previously, we proposed a manual method to find the anatomical views 2C and 4C, in which perpendicular 2D planes from a 3D data set were annotated iteratively [7, 8]. By annotating the apparent apex and two points on the mitral valve ring (anterior/inferior in the approximate 2C and septal/lateral in 4C) a local approximation of the LV long-axis (LAX) is found. The perpendicular plane through the LAX is generated and annotated similarly, generating a better LAX (Figure 1). Convergence to the true 3D LAX usually takes only 4-6 iterations. In this study we propose to replace the manual annotations in 2D cross-sections using the fully automated landmark detection.

2.2. Landmark detection

Instead of detecting the landmark points individually, we try to locate the whole left ventricular region containing those landmarks. The region is defined by a box, in which these landmarks have fixed positions (Figure 2). To do this, we place boxes of different sizes and orientations at different positions in the image. For each box, a classifier determines whether it matches the sought structures; the landmark positions are then defined by the fixed positions within the box. So, the landmark positions are indirectly represented by the position, size, rotation and shear of the box.

Fig. 2. Annotated image of 2-chamber view with a rectangle box, where the landmarks have fixed relative positions, within the box.

We represent the image content of the box using simple and easy to compute Haar wavelet-like features, that have been used in similar applications [4] (Figure 3). The output value at each image position is the difference between the sum of the image pixels in the white section and the sum of image pixels in the black section. The pixels of the resulting filtered images (Figure 4) are the features used in the classifier. As we use multiple feature types with different kernel size, we have many features per pixel position in the box. To compute these features efficiently, we used the integral image approach of Viola and Jones [9].

Fig. 4. Example of (a) an image box and (b) its filtered result using the first Haar feature.

Fig. 5. Map of features used by the third classifier after explicitly excluding invalid region values. Positions correspond to an image box as in fig. 4. The classifier focuses near the anatomical landmark locations (red circles). Colors denote the number of features used at a certain positions.

For the feature calculation in our ultrasound images, we had to deal with missing image information, in regions where the templates exceed the image sector boundaries. Similar to Georgescu et al. [3] we used a mask to exclude invalid pixel values. Additionally, from our set of positive training samples we learned which box regions were generally invalid and we excluded the feature values of these regions from both our training and testing sets (Figure 5).

We train classifiers to distinguish between boxes with the landmarks at positions within an acceptable range (positive examples) and all the rest (negative examples). The classifiers are trained using Adaboost, a machine learning algorithm which combines smaller, weak classifiers into a robust, strong one. This is accomplished by iteratively selecting the best weak classifiers. After each iteration, training examples which are misclassified by the current strong classifier are upweighted, to allow the classifier to adapt more to these misclassified examples during the next iteration. In this study, we use small tree classifiers, each using a small number of Haar features, as the weak classifiers. In initial experiments, we saw that such small tree classifiers performed better than single-feature classifiers. The method works in a coarse to fine approach with more pose parameters and finer steps, using a cascade...
of three classifiers. The first classifier detects the approximate center of the left ventricle (irrespective of size and rotation), the second one detects the apex and mitral valve center (so, approximate LV position, size and rotation, but not shear), and the third one detects all three apex and mitral valve point positions (all pose parameters), with more precision. To train the three classifiers for this cascade, different positive and negative examples are generated from the annotated database, to match the three templates in Figure 6. From each patient, we can generate multiple positive examples, by applying controlled amounts of translation, rotation and shear to the ideal box. That way we model the natural variation of LV size and pose and we have the opportunity to use larger number of examples for the training phase, which is important for a robust classifier.

The cascade is then used for detection. First a rectangular box (of fixed size and orientation) is scanned over the 2D image. The first classifier will evaluate each box and decide whether it contains a centered left ventricle. The “hits” (detected positives, fig. 7) are used to generate new candidates for the second classifier. Around the detected LV centers rotated, scaled and slightly translated boxes are placed. After classifying these with the second classifier, the remaining hits will generate new candidates for the third classifier. This time, the boxes will undergo finer translation, rotation, scale and shear, to detect all landmark points accurately. Each hit from the third classifier gives a final estimate of the landmark positions; the average of all given hits for each patient is taken as the final estimate. To limit our hits we pruned the classified candidates after each stage to exclude spatial outliers.

3. EXPERIMENTAL SETUP

The total method was developed in Matlab (version R2007b, MathWorks Inc.) The evaluation is done in three stages: training the classifiers, tuning them to a desired operating point by ROC analysis (tradeoff between correctly classified boxes and misclassified ones) and finally, testing the detection cascade. 85 routinely acquired 3D patient datasets (Philips Sonos 7500 and iE33) of average clinical quality were manually annotated. 60 patients are used to train the classifiers. The remaining 25 are split into batches of five for cross-validation: each time, 20 are used for the ROC analysis and 5 for evaluation. Detection errors were determined by measuring the 3D Euclidean distance of detected and manual landmarks.

The basic idea of ROC analysis is very simple: for a given trained classifier and a labeled test set, define a set of possible operating points and estimate different type of classifier errors at these points. For training our classifiers we used prTools (http://www prttools org/), a pattern recognition toolbox developed in Matlab. For the ROC analysis we used PRSD Studio (http://prsdstudio.com) from PR Sys Design, which runs in Matlab.

4. RESULTS

We present the detection results in the four and two chamber views. A first impression of the classifiers’ performance is given by the ROC curves (Fig. 8). The curves indicate that our classifiers perform quite well, since we can reject almost all negatives while only losing a small fraction of the true positives. The detection errors are summarized in Table 1. A paired t-test was used for statistical significance (p<0.05).

5. DISCUSSION

We decided to follow a multilevel approach, in which the global position of the left ventricle is found first, and then rotation and scaling are found and further refined, to locate the landmark positions. The method produces a reasonably accurate estimation for the landmark positions. We can detect the points in both the 2C and 4C views (Table 1). The multilevel approach is justified, since the results are improving from stage to stage. The results improved significantly from Step 1 to Step 2. In Step 3, further improvement is achieved (although not always significant) and the individual mitral valve points are detected. Others [6] have argued that such a cascaded approach may be less robust, but we did not experience this here.
Table 1. Detection errors on 2C and 4C view of 25 patients.
* statistically significant difference from previous classifier.
† statistically insignificant difference from previous classifier.

<table>
<thead>
<tr>
<th></th>
<th>2C view</th>
<th>4C view</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LV center</td>
<td>6.5±2.5</td>
<td>7.9±4.1</td>
</tr>
<tr>
<td>Apex</td>
<td>7.2±7.0</td>
<td>8.0±5.0</td>
</tr>
<tr>
<td>Mitral valve center</td>
<td>5.6±5.2</td>
<td>8.2±4.7</td>
</tr>
<tr>
<td><strong>Step 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apex</td>
<td>7.1±6.7</td>
<td>7.9±7.1 †</td>
</tr>
<tr>
<td>Mitral valve center</td>
<td>5.2±2.8†</td>
<td>4.8±2.3 †</td>
</tr>
<tr>
<td>Anterior/ Septal</td>
<td>5.8±3.5</td>
<td>5.6±2.7</td>
</tr>
<tr>
<td>Inferior/ Lateral</td>
<td>4.5±3.1</td>
<td>4.0±2.6</td>
</tr>
</tbody>
</table>

Fig. 8. ROC curves of the second classifier in the 2C and 4C view.

We compare our approach to some published approaches which also identify the sought landmark points, the long axis or the apical views’ orientation angles (Table 2). As already mentioned, Van Stralen’s method [1] uses a combination of the circular Hough transform and dynamic programming. We used this approach to detect the apex and mitral valve center in exactly the same database. This approach does not detect the mitral valve points but only the mitral valve ring, resulting in larger errors than our method (Table 2). Compared to the results given by Orderud et al. [10] we find lower errors for the apex and mitral valve center. Leung et al. [8] outperform our method, but we have to take into consideration that that method aimed at registering markers in stress images, given the markers in the rest image, which is an easier task than detecting the markers. Table 2 also shows comparisons with the manual interobserver and intraobserver variation of the manual 3D marker identification of [7] (following the same protocol as in this study). The fact that our errors are close to those interobserver and intraobserver variabilities is very promising, since these represent the error in our ground truth of manual annotation. It also suggests that our method is accurate enough to replace the manual interaction, although this should be further investigated. Lu’s method results in smaller errors than our method does. Several aspects may play a role: better performance of the steerable features, the structure of the cascade, or their much larger database (326 patients from whom 244 were used for training). However, it might also be a matter of a better ground truth, since in our case the observer errors in the ground truth are such that lower error values are hard to achieve.

Table 2. Comparison with other methods.

<table>
<thead>
<tr>
<th></th>
<th>Apex</th>
<th>Mitral Valve Center</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>7.5±3.3mm</td>
<td>5.0±2.3mm</td>
</tr>
<tr>
<td>Orderud [10]</td>
<td>8.4±3.5mm</td>
<td>3.6±1.8mm</td>
</tr>
<tr>
<td>Lu [5]</td>
<td>4.5±3.5mm</td>
<td>3.6±3.1mm</td>
</tr>
<tr>
<td>Leung [8]</td>
<td>7.6±4.8mm</td>
<td>4.5±2.9mm</td>
</tr>
<tr>
<td>Van Stralen [1]</td>
<td>14.7±6.6mm</td>
<td>8.4±5.7mm</td>
</tr>
<tr>
<td>Interobserver [8]</td>
<td>7.1±2.9mm</td>
<td>3.8±1.3mm</td>
</tr>
<tr>
<td>Intraobserver [8]</td>
<td>5.2±2.0mm</td>
<td>3.3±1.5mm</td>
</tr>
</tbody>
</table>

6. CONCLUSIONS

We showed that our proposed approach can detect the anatomical landmark points accurately. Our errors were comparable to the interobserver and intraobserver variability of the manually indicated landmarks, meaning that our approach is really promising. Additionally, the favorable errors rates show that a small database is sufficient when using artificial variations. The algorithm proved to be accurate and robust enough in both planes, so it could be combined with the already existing manual scheme to extract the views and result in a fully automated method of detecting 3D anatomical landmarks of the left ventricle.

7. REFERENCES