Automatic Estimation of Left Ventricular Dysfunction from Echocardiogram Videos

David Beymer Tanveer Syeda-Mahmood Arnon Amir Fei Wang
IBM Almaden Research Center
650 Harry Rd, San Jose, CA 95120 USA
{beymer, wangfe}@us.ibm.com, {stf, arnon}@almaden.ibm.com

Scott Adelman, M.D.
Kaiser Permanente
San Rafael, CA 94903
scott.adelman@kp.org

Abstract

Echocardiography is often used to diagnose cardiac diseases related to regional and valvular motion abnormalities. Due to the low resolution of the imaging modality, the choice of viewpoint and mode, and the experience of the sonographers, there is a large variance in the estimation of important diagnostic measurements such as ejection fraction. In this paper, we develop an automatic algorithm to estimate diagnostic measurements from raw echocardiogram video sequences. Specifically, we locate and track the left ventricular region over a heart cycle using active shape models. We also present efficient ventricular localization in video sequences by automatically detecting and propagating echocardiographer annotations. Results on a large database of cardiac echo videos demonstrate the use of our method for the prediction of left ventricular dysfunction.

1. Introduction

Echocardiography is often used to diagnose cardiac diseases related to regional and valvular motion abnormalities. It provides images of cardiac structures and their movements from which echocardiographers extract important measurements to estimate heart performance such as ejection fraction and ventricular volume at specific points in the heart cycle.

However, due to the low resolution of the imaging modality, the choice of viewpoint and mode, and the subjective judgment of sonographers, many of these measurements lack precision in diagnosis. Table 1 illustrates this lack of precision through sample ejection fraction (EF) estimates computed from various sources for a few patients. Ejection fraction measures the ratio of stroke volume to the end-diastolic ventricular volume [7]. In Table 1, the columns indicate ejection fraction measurement made in four-chamber view, two-chamber view, M-mode, and the bi-plane method of disks[7]. The corrected measurement recorded by the cardiologist in the actual echocardiogram report is shown in the sixth column. Thus we can see considerable variation in the manual estimates of this measurement.

Such variance can lead to diagnosis errors as physicians continue to make judgments based on evidence from a single patient's data. The goal of our research has been to mine large pre-diagnosed patient exam data collections to capture the statistical correlation between the diagnosis and the variance in measurements, for more informed decision support. To enable such mining for cardiac echo video datasets, it should be possible to automatically derive various diagnostic measurements from raw video sequences. This can be a challenging problem, requiring segmentation of the video into different imaging modes (M-mode, Doppler, video cineloops, etc), the recognition of echocardiogaphic viewpoints (four-chamber views versus two-chamber views, for example), and the isolation of cardiac regions within each viewpoint (see Figure 1, left). While echocardiographic viewpoint recognition has been addressed lately by several researchers[13, 3], completely automatic isolation of cardiac regions and measurement of diagnostic parameters from such regions, is still not well-explored.

In this paper we address the automatic estimation of diagnostic measurements such as ejection fraction from raw echocardiograms. Specifically, we process a raw echochardiogram video sequence to find the relevant views depicting the left ventricular region. We then track the changes in appearance of the ventricular region through the heart cycle using active shape models (ASMs). Active shape models are nonrigid shape models that can capture shape and textural information within a region. As described in [3], they are generated from training across a large collection of sample region appearances under various disease conditions. Due to the flexibility offered in these models, they can cause false positives and confusion with other heart chambers. To reduce the false positives as well as to avoid the computational expense of localizing the shape models within each echo video frame, we

978-1-4244-3993-5/09/$25.00 ©2009 IEEE

Table 1. Illustration of variance in ejection fraction estimates by manual measurement.

S.No.	Apical 4-Chamber	Apical 2-Chamber	Biplane Method-of-disk	M-mode	Report Corrected	Auto-estimated
1.	59.1	60.9	59.9	67.3	55-60	57.098
2.	64.7	66.9	65.7	75.2	60-65	61.088
3.	82.4	86.4	84.6	61.2	60-65	65.377
4.	47	57.3	53.1	54.8	45-50	53.6
5.	47.8	52.4	48.7	36.9	40-45	49.848
6.	46.4	60.6	51.8	55.2	50-55	62.423
7.	63.2	64.2	61.4	46.7	60-65	64.934

restrict the analysis to those clips/sequences that are adjacent to a manually annotated frame. These annotations typically outline a heart region and make measurements within these regions as shown in Figure 3a. We automatically extract such regions, and derive templates for detecting candidate left ventricular regions within adjacent video clips using correlation. Active shape models are then applied to accurately localize the ventricle and to track it over a heart cycle. This allows us to estimate not only conventional measurements such as ejection fraction, but also new diagnostic measurements that show the variation of ventricular volume continuously through the heart cycle. We show that our method computes ejection fraction estimates that are within the range of estimates produced by manual measurement.

The rest of the paper describes our method in detail. In Section 2, we review relevant literature pertaining to the automatic computation of diagnostic measurements. In Section 3.1, we describe active shape models for modeling and tracking ventricular regions. In Section 3.2, we describe left ventricle (LV) localization using annotations. In Section 3.3, we discuss the measurement of ejection fraction within the regions modeled by ASMs. In Section 4, we present results of comparison of LV assessments produced by our method with those collected by echocardiographers on a large collection of cardiac echo videos.

2. Related Work

The automatic estimation of diagnostic measurements such as ejection fraction has been attempted by a number of researchers [2, 14]. Some of this work has been done using cardiac MRI or SPECT (as against echocardiograms) [9, 15, 16, 19]. Different algorithms for EF estimation have been compared as well [11]. The visual estimation of EF was studied across various echocardiographers in [1]. State-of-the-art algorithms can now be found in echo machines to automatically localize the left ventricle and estimate the ejection fraction [4, 12, 18]. For example, in [4], an LV region was localized using template matching with a large collection of previously acquired templates. A shape model for the region and its contour was then created using a database of normal and abnormal LV shapes as a guide [8]. In [10], a semi-automatic segmentation algorithm was proposed to outline the endocardium in apical views of echocardiograms. More recently, an active contour based approach was used to segment the LV region in [17].

Our work differs from those in the literature in several respects. Prior work attempts to estimate diagnostic measures within echo frames where the viewpoint is already known. Our work, on the other hand, does not assume that the viewpoint is known and works with an unannotated video sequence. Secondly, we use a more flexible model of shape, appearance and texture derived through feature correspondence between candidate training region samples, to build a model for the LV region. This is unlike other approaches that are based on affine deformations of regions from standard stored templates.

3. System for estimating LV dysfunction

Our overall approach to automatic assessment of LV dysfunction is summarized in Figure 1. The input to our system is the complete set of videos and still images in an echo study. The raw sequence depicts various imaging modes such as CW Doppler, M-mode, as well as different viewpoints such as A4C views and A2C views. From this raw sequence, we first try to detect sonographer annotations of the left ventricle. If present, we leverage the annotation to find nearby LV cine loops in the raw sequence and to initialize an ASM fit in these cine loops (upper pipeline in Figure 1). If sonographer annotations are not present, then we use the approach described in [3] to segment the modes and to recognize the A4C views. An LV detector initializes ASM fits in A4C cine loops in this case (lower pipeline in Figure 1). Finally, once the LV ASM models are fit, important diagnostic measurements are extracted as explained in Section 3.3.

3.1. Modeling LV region using shape models

To extract the left ventricular region and to track its changes within the heart cycle, we model the LV region using an active shape model. Our approach to using active shape models (ASMs) for locating and tracking the LV region is similar to the one used earlier for recognizing viewpoints in echo videos[3]. ASMs are an appearance-based modeling approach that are particularly adept at handling

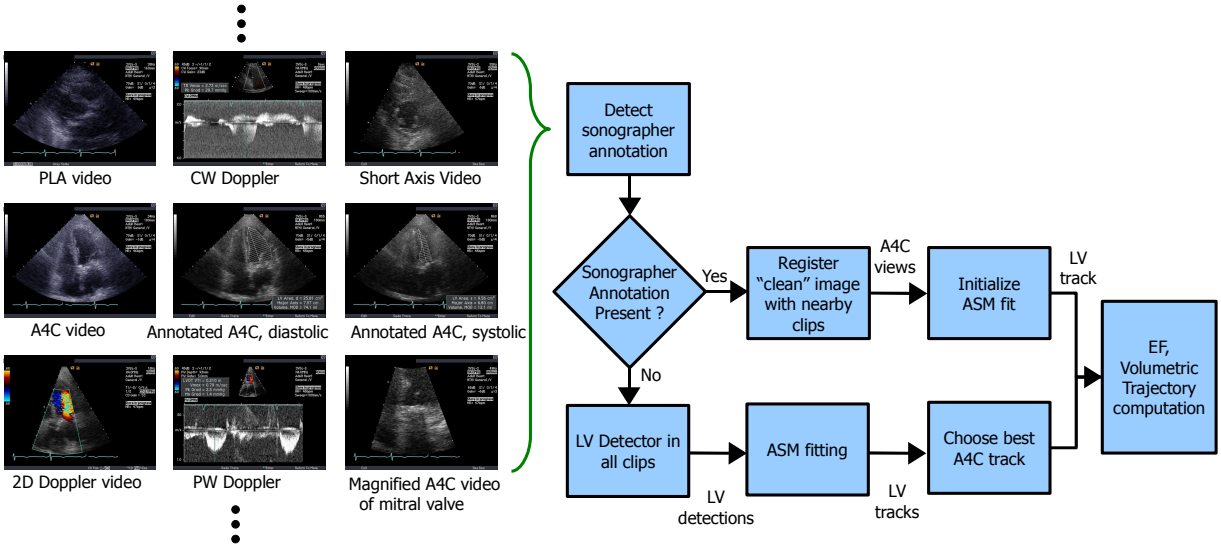


Figure 1. Flowchart of our system for evaluating LV dysfunction.

nonrigid classes of objects [6]. ASMs divide the appearance of an object class into a shape vector $\mathbf{s}$ and texture vector $\mathbf{t}$. Shape $\mathbf{s}$ is a concatenation of the (x, y) coordinates of n feature points, and, texture $\mathbf{t}$ is a concatenation of pixel values from patches centered on each of the feature points. The mean shape and texture vectors obtained using principle component analysis (PCA) are used to form the active shape model as:

$$\begin{aligned} \mathbf{s} &= S\,\mathbf{a} + \bar{\mathbf{s}} \qquad S = \begin{bmatrix} | & | & & | \\ \mathbf{e}_1^s & \mathbf{e}_2^s & \dots & \mathbf{e}_p^s \\ | & | & & | \end{bmatrix} \\ \mathbf{t} &= T\,\mathbf{b} + \bar{\mathbf{t}} \qquad T = \begin{bmatrix} | & | & & | \\ \mathbf{e}_1^t & \mathbf{e}_2^t & \dots & \mathbf{e}_q^t \\ | & | & & | \end{bmatrix} \end{aligned} \tag{1}$$

where p eigenshapes $\mathbf{e}_1^s, \mathbf{e}_2^s, \dots, \mathbf{e}_p^s$ and q eigentextures $\mathbf{e}_1^t, \mathbf{e}_2^t, \dots, \mathbf{e}_q^t$ are retained in the PCA. The p-dimensional vector $\mathbf{a}$ and the q-dimensional vector $\mathbf{b}$ are the low dimensional representations of shape and texture (from [3]).

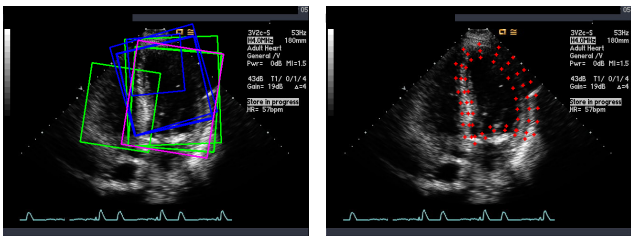
To model the LV region using active shape models, we collected a number of training images covering different patients, diseases, and time offset within the cycle for a known viewpoint (in our case, A4C views). The LV region was then represented by the shape and texture information for a set of feature points in the LV outer walls and LV endocardium. While the feature points are manually isolated during training stage, they are automatically identified during matching.

Assuming a candidate LV region can be isolated in an echo video frame, fitting an ASM model to that region involves finding a similarity transform Γ^{sim} and vectors $\mathbf{a}$ and $\mathbf{b}$ that align well with the content in the region. Using an analysis-by-synthesis approach, we alternately update the shape and texture models as described in [3]. Once fitting has converged (see Fig. 2(b) for an example), the fit is evaluated using

$$\text{fit}(\mathbf{a}, \mathbf{b}, \Gamma^{sim}) = \mathbf{a}^T \Sigma_{\text{shp}}^{-1} \mathbf{a} + \mathbf{b}^T \Sigma_{\text{tex}}^{-1} \mathbf{b} + 2R^2/\lambda_{\text{tex}}^{q+1}, \tag{2}$$

where $R = \|\mathbf{t} - TT^T\mathbf{t}\|$, $\mathbf{t} = I(\Gamma^{\text{sim}}(x, y))$, $\lambda_{\text{tex}}^{q+1}$ is the $(q+1)^{\text{th}}$ texture eigenvalue, and Σ_{shp} and Σ_{tex} are diagonal matrices with PCA eigenvalues (see [3, 5]).

In the absence of any annotations to guide the selection of the LV region, we first apply an LV detector. We use a "distance-to-eigenspace" approach based on a set of training examples with annotated LV shapes. In an offline training step, the mean LV shape is computed and all the examples warped to a "shape free" representation at mean shape. PCA is applied to this normalized set, and the eigenimages with larger eigenvalues retained. At run-time, the LV detector first tests different translations for the eigenimage model, sliding a window over different (x, y) offsets and computing the distance from the input window to the eigenspace. Offsets with a distance below a threshold are further examined, but generalizing the transform from translation to a full similarity transform. At a hypothesized window seeded by the initial distance-to-eigenspace, we iterate between PCA projection, reconstruction, and similarity transform update using motion templates. This generates a set of LV detection boxes as shown in Fig. 2(a). ASM fitting is used to track the LV for each hypothesis through the cardiac cycle, and the ASM track with best fit (eqn 2) averaged over the cycle is chosen as the A4C track for the clip. To process an entire echo study, we are presented with



a) LV detector b) ASM model fit

Figure 2. (a) LV detection results and (b) ASM fitting results.

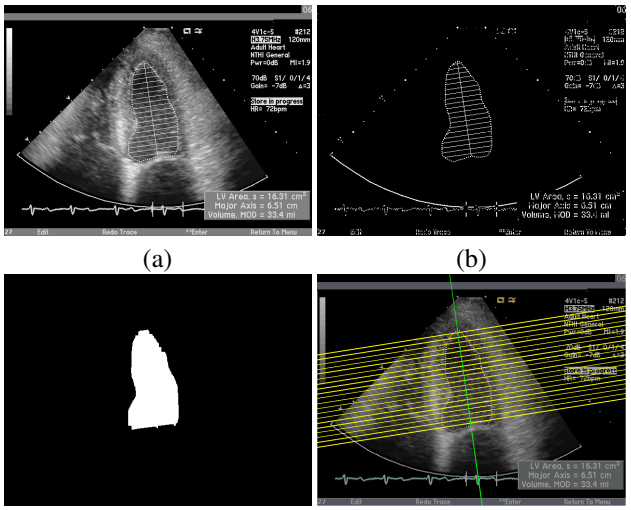
a number of clips, most of which are not A4C. LV detection and A4C tracking are applied to all non-Doppler video clips that are not zoomed in (e.g. zoomed in on a valve). The clip with the best A4C fit is taken as the most representative A4C view and subsequently used for computing ventricular volume and EF.

3.2. Processing echocardiographer annotations

During an echo study, a sonographer often annotates regions (left ventricle, right atrium, etc) in an A4C or A2C viewpoint "still image" as part of using the method-of-disks technique for estimating the regional volume. An example is shown in Figure 3(a) where the LV region is marked by the major axis and is divided into orthogonal parallel lines for measuring the ventricular volume. This annotation is captured in the resulting echo video sequence for the patient as a still image intermixed with the cine loops. We now describe the automatic isolation of such regions to serve as a prior for faster localization of LV regions for ASM fitting within A4C cine loops.

The first temptation is to search for the straight lines, using edge detection and Hough transform. However, the noise in echo images, the very thin, short lines and high intensity echo regions in the background degrade the robustness of both edge detection and Hough transform. To achieve a robust detection, a different approach is needed.

Our approach is composed of a number of steps, as illustrated in Figs. 3-8. First, pixels along thin lines are detected using image processing in local neighborhoods (Figure 3(b)). Next we use mathematical morphology to remove most false positives, grow regions, and detect the annotated region, as shown in Figure 3(c). The region mask is then used for robust and accurate lines detection (Figure 3(d)), and OCR of image text is used to label the detected heart chamber (LV vs. LA vs. RA, see Figure 6). Finally, a template from a "cleaned" LV detection is used to locate the LV in nearby A4C cine loops (Figure 7) and initialize an ASM fit (Figure 8).



(a) (b) (c) (d)

Figure 3. a) Original image, showing annotated LV at end of diastolic period, b) difference image, I_l, c) detected region, and, d) detected region and lines, superimposed on the original image.

3.2.1 Detecting thin line components

To isolate the annotated region, we use morphology operations to detect local pieces of thin bright lines and fill-in the intervening regions. For this, we observe that the lines in the annotated region are single-pixel wide, and brighter than their background. To extract these lines, we first form a background image by applying a median filter. Let I, I_m denote the original frame and the $5x5$ Median filter of it, respectively. The background image, I_b, denotes the frame where all the locally bright thin features are removed.

$$I_b(i,j) = \begin{cases} I_m(i,j) & I(i,j) > T_b \; and \\ & I(i,j) - I_m(i,j) > T_c \\ I(i,j) & otherwise \end{cases} \quad (3)$$

where T_b and T_c are brightness and contrast thresholds. Values of $T_b = 180$ and $T_c = 30$ were found to be sufficient to separate the true line pixels from the background.

The foreground image I_f, or complement of I_b, is found by thresholding the difference of I_b and the original image: $I_f(i,j) = 1$ if $I(i,j) - I_b(i,j) > T_c$, and, otherwise, $I_f(i,j) = 0$. This captures thin lines and any other thin and bright features in the image.

Next we will use morphology operations to isolate thin lines in the foreground image. Let $\ominus$,$\oplus$, $\bullet$, $\circ$, $\setminus$ and $\otimes$ denote image erosion, dilation, opening, closing, set subtraction and xor operations, respectively. Lines with orientation lower than 30° (or above 60°) are detected as regions containing horizontal (vertical) elements but no vertical (horizontal) ones, that is $I_1 = (I_f \circ S_a) \otimes (I_f \circ S_b)$ where S_a and S_b are 2×2 pixel structuring elements, shown in Figure 4. Similarly, thin lines with slope of roughly $30 - 60°$ (nearly

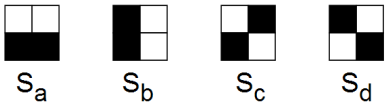


Figure 4. Structuring elements, used for thin-lines detection.

diagonal) are detected as regions containing diagonal elements of one diagonal direction and excluding elements of the other diagonal direction: $I_2 = (I_f \circ S_c) \otimes (I_f \circ S_d)$. Finally, $I_l = I_1 + I_2$ includes thin lines of any orientation, as shown in Figure 3(b). Note that any non-thin regions are filtered out through this process, as they are not significantly affected (eroded) by opening with any of the four small structuring elements, S_a, S_b, S_c and S_d. This method of detecting the lines is more robust than using a conventional edge detector or the Hough transform due to their short length and small thickness.

Next, to thicken the lines and connect them into a single region we apply a few dilation and close operations. Other image structures, such as text, may also form into regions. We compute connected components and select the largest component as the annotated region of interest M_{roi}. The region of interest (ROI) mask image so detected for the still frame of Figure 3(a) is shown in Figure 3(c).

The eigenvectors of the annotated region could be used to denote its orientation. However, due to the irregular shape of the left ventricle in diseased cases, it is preferable to rely on the orientation indicated by the echocardiographer. We now describe a robust method to estimate this orientation.

3.2.2 Region-based normalized Hough transform

For the orientation estimation, we restrict the analysis to the annotated region as $I_{froi} = I_f \cdot M_{roi}$, where $\cdot$ is a pixel-wise multiplication. A standard Hough transform, $H(\theta, \rho) = Hough(I_{froi})$ may detect some of the lines. However, when the region is very narrow the lines are short, and the corresponding peaks in Hough space are too low to be reliably detected, leading to many misses and false positives. To achieve a more robust line-detection within the region of interest, we define the *region-based normalized Hough transform*, as

$$\hat{H}(\theta, \rho) = H(\theta, \rho)/(H_{ref}(\theta, \rho) + H_c) \qquad (4)$$

The term H_{ref} is a *reference* Hough transform, of the mask itself. That is, $H_{ref}(\theta, \rho) = Hough(M_{roi})$, where each pixel in the mask area is counted as an edge pixel for the Hough transform. Hence each cell in $H_{ref}(\theta, \rho)$ contains the number of pixels (or length) expected in a corresponding line of the masked image. The region-based normalized Hough transform is thus a ratio between two

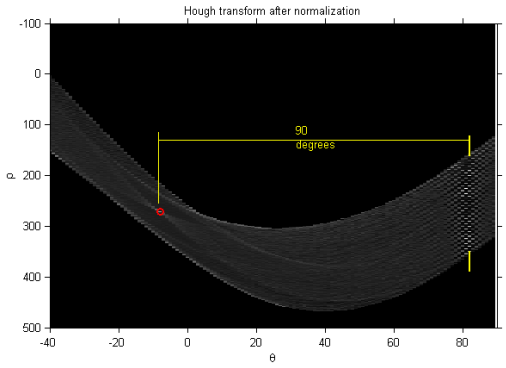


Figure 5. Region-based normalized Hough transform, marked with a line at θ_g and a circle around the point corresponding to the major axis.

Hough transforms, $\hat{H}(\theta, \rho) = H(\theta, \rho)/(H_{ref}(\theta, \rho) + H_c)$ where $H_c = 10$ is a smoothing factor, to eliminate false detection of very short "lines" (a few pixels long), where the mask is narrow. By definition, for any ρ and θ, the region-based normalized transform is bounded $0 \leq \hat{H}(\theta, \rho) \leq L/(L + H_c) \leq 1.0$, where L is the total line length under the mask M. It approaches 1.0 for ideal lines of long lengths. We refer to it as a normalization in Hough space - a useful property as our lines of interest greatly vary in length, according to the shape of the region.

3.2.3 Extracting line parameters

A structure made of a group of parallel lines has a special appearance in Hough space. It shows as a set of peaks, sharing the same $\theta_i = \theta_g$ where θ_g denotes the group slope, and having different ρ_i values. The line detection is made very robust by simultaneously detecting the parallel structure. Let $S(\theta) = \sum_\rho (C(\theta, \rho)$ where $C(\theta, \rho) = 1$ if and only if $\hat{H}(\theta, \rho) > T_H$, where $T_H = 0.45$ is the single-line detection threshold. We first find $\theta_g = arg\ max_\theta\{S(\theta)\}$. We then look at the two adjacent θ-s, to compensate for quantization errors, and detect the lines one by one by finding the peaks for rho_i.

Lastly, we detect the major axis as orthogonal to the parallel lines. We limit the search for it in the Hough space to $\theta \pm \pi/2$. This is illustrated in Figure 5. The detected lines are shown in Figure 3(d).

3.2.4 Identifying the annotated chambers

Since the echocardiographers annotate other regions beside the left ventricle, the identity of the isolated region needs to be established. Fortunately, the identity can be inferred from the measurements which are included in the captured sequence. A popup box, usually in the lower right corner of

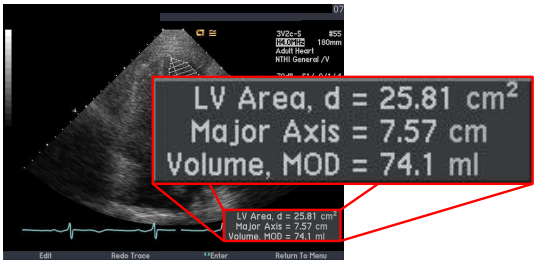


Figure 6. Popup text on annotated frames describes chamber (LV, RV, RA), position in cycle (diastolic, systolic), and, indirectly, viewpoint (A4C, A2C).

the image, displays summary information about the chamber region, including its area, volume, and major axis as shown in Fig. 6. To extract and recognize the text in such regions, we use an optical character recognition (OCR) engine (Tesseract). From the recognized text, we can infer the name of the chamber (LV, LA, RA) as well as the position within the cycle (diastolic, systolic), which will be useful in locating the LV region in adjacent clips as described next.

3.2.5 Propagation of annotations for LV localization

A typical echo study sequence consists of hundreds of cine loops intermixed with still frames depicting annotated regions. In most practical cases, it is sufficient to attempt LV localization within a neighborhood $[i-10, i+10]$ of an annotated still frame i (see Fig. 7). To isolate candidate frames for locating the LV region in these sequences, we use the background image of Equation 3 as a template. By moving the template over the frames of the neighboring cineloops, we try to find the best similarity transform-based alignment of the template with the candidate region. The video frames where the normalized correlation coefficient of the template is over a threshold, are retained as candidate frames to localize the LV region. Figure 7 illustrates the registration of annotated images with frames of neighboring clips.

The registration step has set up a similarity transform between clips i and j based on image-image matching of grey levels. We now use the same transform to map the annotation mask in Fig. 8(b) to (d), effectively transferring the annotation to the video clip. Next, the boundary of the transferred mask (d) is computed, as this corresponds to the boundary of the endocardium with the blood pool. Once the mitral valve attachment points are estimated (blue/red dots in (d)), we bring the endocardial boundary into correspondence with the same features in the ASM model. The ASM model can be initialized from these correspondences, producing initial estimates for a normalizing similarity transform and nonrigid shape (Fig. 8(e)). The final result of ASM fitting (see section 3.1) is shown in Fig. 8(f).

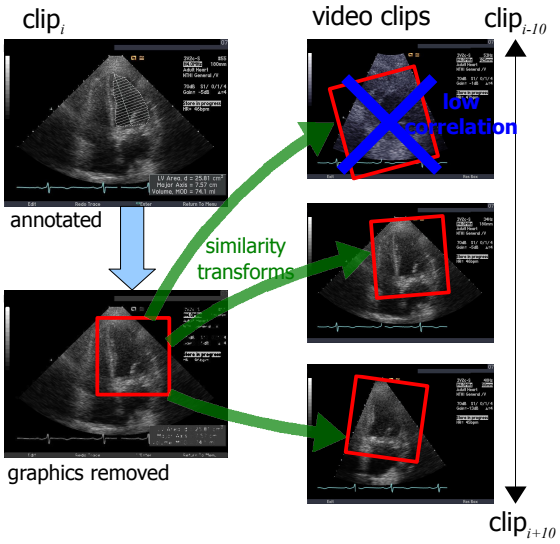


Figure 7. To transfer the view label and map image of an annotation, we register a clean template of the chamber against nearby video clips using image-image registration techniques.

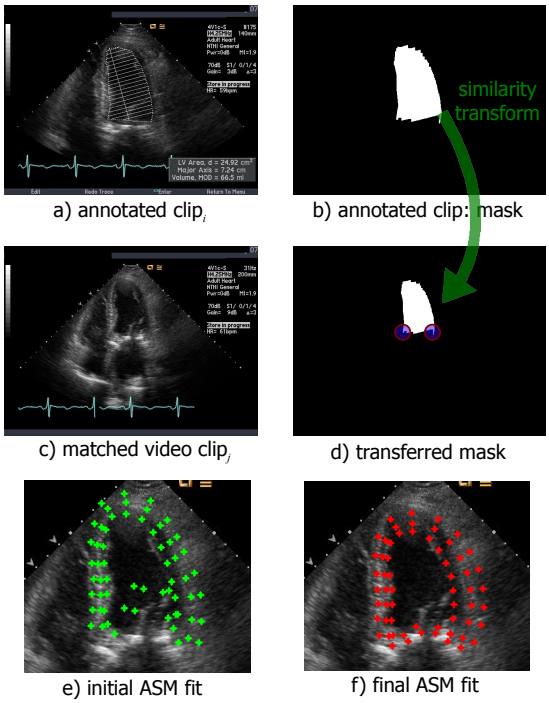


Figure 8. Using the annotation in clip$_i$ to initialize an ASM fit in clip$_j$. Please refer to the main text for details.

3.3. Automatic assessment of LV function

LV function is a measure of the pumping ability of the heart. To estimate the volume of blood being pumped on

each heart cycle, sonographers look at the boundary between the blood pool, which appears dark in the echo image, and the endocardium, the inner boundary of the cardiac muscle. While in 2D echo, this chamber region is just an area, there are established techniques, notably the method of disks, for estimating the 3D volume of the chamber[7]. In 2D echo, it is common to measure volume at two points in the cardiac cycle: end diastole (max volume) and end systole (min volume). EF, or ejection fraction, looks at the fraction of the volume that is pumped out on every cycle

$$EF = \frac{EDV - ESV}{EDV}$$

where EDV is end diastolic volume and ESV is end systolic volume. The normal range for EF is in the 65-100% range, while an EF below 30% indicates severe LV dysfunction.

With the automatic isolation of annotated regions, it could be argued that there is sufficient information to automatically compute the ejection fraction. However, doing so will again have consistency problems as indicated earlier in Table 1. Further, continuous estimation of chamber volume within the heart cycle would not be possible. Since our ASM model uses features derived from the endocardium and the outer wall boundary of the LV, we can now automatically apply the method of disks to the region enclosed by the inner ASM contour. As shown in Fig. 9, the major axis of the region is first connected from the apex of the heart to the midpoint of the segment connecting the mitral valve wall attachments. Then the LV chamber is divided into 20 disk segments, with each segment perpendicular to the major axis, and the estimation in 3D done by sweeping out each disk as a surface of revolution. The estimated volume is

$$\text{LV Volume} = \frac{\pi}{4} \sum_{i=1}^{20} d_i^2 \times \frac{L}{20}$$

where d_i is the diameter of the ith disk, and L is the length of the major axis. For the disk diameters and L, we can map from pixels to cm using an automated procedure for finding the calibration markers on the side of the echo sectors. Fig. 9(top) shows the major axis and disks generated from an ASM LV track, and the figure bottom shows a volumetric trajectory within the heart cycle by instantaneous estimation of chamber volume. The computed EF from this ASM LV track is 59%, compared to a physician report range of 65-70%, which is within 6% of the physician range.

4. Results

We now report on the results of automatic computation of ejection fraction from whole echocardiogram sequences. A large database of 1771 cardiac echo video studies were collected from 1178 cardiac patients from a large hospital

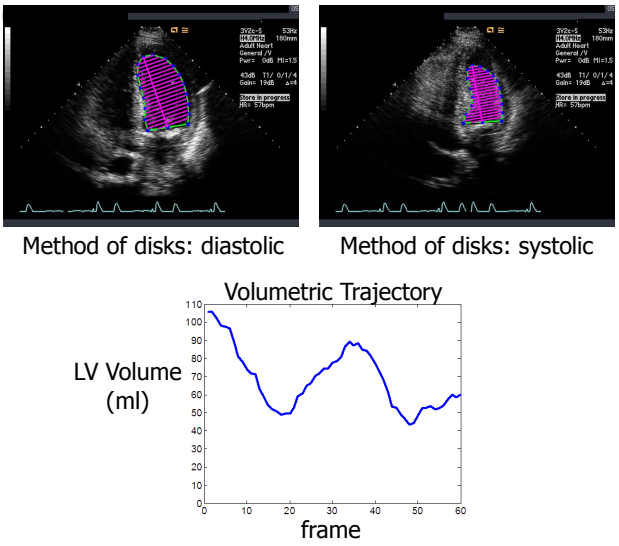


Figure 9. (top) Method of disks applied on an automatic A4C LV track for end diastolic and end systolic frames, and (bottom) a complete volumetric trajectory.

network in our area. Each echo study on the average, had between 50-100 clips depicting multiple viewpoints. Each clip was 30-60 video frames long giving rise to over 5 million image frames.

4.1. Annotation region detection accuracy

We first evaluated the accuracy of automatic extraction of annotation regions from the videos. Our annotation algorithm was found to be very robust and failed on detecting the region in only 0.1% of the dataset. Almost all these cases occurred when the textual measurements were directly overlaid on the annotated regions, occluding the parallel-line groups and interfering with the morphological operations.

4.2. Improvement of performance using annotation

Next, we compared the search performance during ASM model fitting with and without the propagation of annotations. The generation of ASM fit candidates using ASM models for A4C views on the entire data set of 5 million frames took a week of processing. In comparison, using the annotated regions, the corresponding processing time was only a few hours. In addition, using the location prior supplied, automatically computed EF estimates agreed with physician-corrected estimates in reports in 9.4% more cases. The EF estimates computed for a subset of patients of Table 1 are shown in the last column of this table. As can be seen, the estimates produced are within the range of manual estimates.

4.3. Estimation of LV dysfunction

Finally, we tested the performance of prediction of LV dysfunction using our automatic estimation method based on ASMs in comparison to the manually determined values. Specifically, the LV dysfunction was divided into three ranges, namely, normal-mild, moderate and severe corresponding to the ejection fraction ranges of $(L_1, U_1) = (0, 0.3)$, $(L_2, U_2) = (0.3, 0.6)$, and $(L_3, U_3) = (0.6, 1.0)$, respectively. Since the manual measurements show considerable variation, an interval was made using the minimum and maximum values of manually computed EF estimates and was denoted as the reference interval (L_r, U_r). The overlap of the reference interval with the LV dysfunction ranges above was noted as $O(r, j) = min(U_j, U_r) - max(L_j, L_r)$, and the patient was classified as having mild, moderate or severe dysfunction based on the range with the maximum overlap with the reference interval, $O(r, j)_{max} = \max_j O(r, j)$. The prediction accuracy for our automatically computed estimate f_p for a patient is then determined as

$$\text{prediction accuracy} = \frac{|f_p | f_p \in O(r, j)_{max}|}{N} \quad (5)$$

Of the 1771 cases, Automatic EF estimation gave no output for 167 cases or 9.4%. For the prediction of severe LV dysfunction (EF < 0.3), the algorithm agreed with at least one human interpretation (by any mode) in 256 of the 256 cases or 100% For the prediction of mild to moderate LV dysfunction (EF, 0.3-0.6), the algorithm agreed with at least one human interpretation in 381 out of 722 cases or 53% For the prediction of normal LV function (EF > 0.6), the algorithm agreed with at least one human interpretation in 393 out of 626 cases or 63%.

5. Conclusions

This paper presents an automatic method for unsupervised computation of ejection fraction from raw echocardiogram videos. In particular, active shape models were used to capture the shape and textural information within LV region. The localization of the LV region was made efficient by propagating the annotation of echocardiographers in nearby still frames. Results on a large database of cardiac echo videos demonstrate the use of our method for prediction of left ventricular dysfunction.

References

[1] O. Akinboboye et al. Visual estimation of ejection fraction by two-dimensional echocardiography: The learning curve. *Clinical Cardiology*, 18(12):726–729, 1995. 2

[2] U. Barcaro, D. Moroni, and O. Salvetti. Automatic computation of left ventricle ejection fraction from dynamic ultrasound images. 18:351–358, 2008. 2

[3] D. Beymer, T. Syeda-Mahmood, and F. Wang. Exploiting spatio-temporal information for view recognition in cardiac echo videos. In *MMBIA*, pages 1–8, 2008. 1, 2, 3

[4] M. Cannesson et al. A novel two-dimensional echocardiographic image analysis system using artificial intelligence-learned pattern recognition for rapid automated ejection fraction. 49:217–226, 2007. 2

[5] T. Cootes and C. Taylor. Using grey-level models to improve active shape model search. In *International Conference on Pattern Recognition*, volume 1, pages 63–67, 1994. 3

[6] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Comput. Vis. Image Underst.*, 61(1):38–59, 1995. 3

[7] H. Feigenbaum, W. F. Armstrong, and T. Ryan. *Echocardiography, Sixth Ed.* Lippincott Williams & Wilkins, 2005. 1, 7

[8] B. Georgescu, X. S. Zhou, D. Comaniciu, and A. Gupta. Database-guided segmentation of anatomical structures with complex appearance. *IEEE CVPR*, 2:429–436, 2005. 2

[9] G. Germano et al. Automatic quantification of ejection fraction from gated myocardial perfusion SPECT. 36:2138–2147, 1995. 2

[10] M.-P. Jolly. Assisted ejection fraction in b-mode and contrast echocardiography. In *ISBI*, pages 97–100, 2006. 2

[11] M. Kupinski et al. Comparing cardiac ejection fraction estimation algorithms without a gold standard. 13:329–337, 2006. 2

[12] E. Maret et al. Computer-assisted determination of left ventricular endocardial borders reduces variability in the echocardiographic assessment of ejection fraction. 6:55, 2008. 2

[13] J. Park, S. Zhou, C. Simopoulos, J. Otsuki, and D. Comaniciu. Automatic cardiac view classification of echocardiogram. In *International Conference on Computer Vision*, pages 1–8, 2007. 1

[14] J. Park, S. K. Zhou, J. Jackson, and D. Comaniciu. Automatic mitral valve inflow measurements from doppler echocardiography. *Proc. MICCAI 2008*, pages 983–990, 2008. 2

[15] S. Sinha, R. Mather, U. Sinha, J. Goldin, G. Fonarow, and H. Yoon. Estimation of the left ventricular ejection fraction using a novel multiphase, dark-blood, breath-hold mr imaging technique. 169:101–112, 1997. 2

[16] M. B. Stegmann and D. Pedersen. Bi-temporal 3D active appearance models with applications to unsupervised ejection fraction estimation. In *International Symposium on Medical Imaging 2005, San Diego, CA,*, volume 5747, pages 336–350. SPIE, feb 2005. 2

[17] J. Sun et al. Automated echocardiographic quantification of left ventricular volumes and ejection fraction: Validation in the intensive care setting. 8:29–36, 1995. 2

[18] SYNGO. SYNGO Auto EF, Siemens Medical Solutions, 2006. 2

[19] P. Thunberg, K. Emilsson, P. Rask, and A. Kahari. Separating the left cardiac ventricle from the atrium in short axis mr images using the equation of the atrioventricular plane. 28:221–228, 2008. 2