

Inferring Transducer Viewpoints in Cardiac Echo Videos

D Beymer, T Syeda-Mahmood, F Wang

IBM Almaden Research Center, San Jose, CA, USA

Abstract

During a 2D echo exam, the transducer position is varied to elicit important information about the heart function and its anatomy. Knowledge of the transducer viewpoint is important in automatic cardiac echo interpretation to understand the regions being depicted as well as in the quantification of their attributes. In this paper, we address the problem of inferring the transducer viewpoint from the spatio-temporal information in cardiac echo videos. Unlike previous approaches, we exploit motion of the heart within a cardiac cycle in addition to spatial information to discriminate between viewpoints. Specifically, we use an active shape model (ASM) to model cardiac appearance in an echo frame, and cardiac motion is modeled using eigen-motions of ASM feature tracks. We test our method on 2 data sets of patients with various cardiac diseases, and we report comparison with re-implementations of two state-of-the-art view recognition methods.

1. Introduction

2D Echocardiography is an important diagnostic aid for morphological and functional assessment of the heart. Much of the practice of echocardiography requires manual intervention in both imaging and in interpretation. In particular, the transducer position is varied during an echo exam to capture different anatomical sections of the heart from different viewpoints. Methods for automatic interpretation of echocardiograms are now becoming increasingly available. However, due to the large difference in appearance of the cardiac anatomy under different transducer positions, many of these methods find it difficult to make accurate cardiac assessment without the knowledge of the transducer position. Thus automatic inference of transducer viewpoint is important in automatic cardiac echo interpretation for a better understanding of the regions being depicted as well as in the quantification of their attributes. This in turn will enable better computer-aided diagnosis and decision support in future.

Computer vision researchers have worked on the problem of cardiac view recognition, especially for the popular views such as apical four chamber, parasternal long axis, short axis view, etc. [1–5]. In the parts-based

approach of [1], heart chambers are modeled using a relational graph and Markov Random Fields for each viewpoint. In [5], the standard views were represented by templates obtained by applying multiresolution spline filters to intensity images. After elastically matching a new view to the templates, the deformation energy and similarity match are fed as features to a linear discriminant classifier. In [3] and [4], view recognition is cast as a problem of multi-class object detection. After filtering the image intensities with Haar wavelets, a multi-class classifier using ML Boosting is applied, with one class per viewpoint.

Our spatiotemporal modeling approach presented here utilizes the machinery of active shape models (ASM) [6] and active appearance models (AAM) [7]. ASMs and AAMs are effective statistical tools for modeling the appearance and texture of nonrigid objects, and have been extensively applied in computer vision. They have also been previously used for echocardiogram segmentation and tracking [8, 9]. A variant of AAMs, Active Appearance Motion Models (AAMMs) [10] is most similar to our work. In AAMMs, shape and texture are represented jointly across the entire cardiac cycle, and motion is captured in the shape vector. While their shape/motion modeling is used to constrain tracking, we will use motion eigenanalysis as a feature for viewpoint classification.

In this paper, we address the problem of inferring the transducer viewpoint from the spatiotemporal information in cardiac echo videos. Unlike previous approaches, we exploit motion of the heart within a cardiac cycle in addition to spatial information to discriminate between viewpoints. Specifically, a statistical spatiotemporal model of cardiac appearance called the Active Shape Model (ASM) is constructed for each echo viewpoint. To represent the characteristic heart motion over the cardiac cycle, we track ASM features over the cycle and construct linear PCA motion models for each echo viewpoint. Given a video of an unknown echo viewpoint, we fit the combined ASM and motion models from each viewpoint to the input video, choosing the model with the best fitting parameters.

2. Method

Our approach to view recognition is based on building spatial and temporal models for known viewpoints from sample learning data and using a matching algorithm for

resolving the best matching model. The spatial and textural information across frames of the echo sequences of a training set are used to build viewpoint-specific active shape models. Next, the ASM models are tracked within the training sequences to generate motion models for each viewpoint class. To determine the viewpoint of an unknown single heart-cycle echo sequence, we fit a candidate ASM to each of the frames in the sequence to create a tracked ASM sequence. Motion information derived by tracking ASMs through a heart cycle is then projected into the eigen-motion feature space of the viewpoint class. The combined fit of the ASMs and the motion models are then used to evaluate the overall spatio-temporal model fit. The best viewpoint class is then determined by a winner-take-all approach.

2.1. Appearance modeling

Using the paradigm of active shape models [6], we represent shape information in an echo frame by a set of feature points obtained from important anatomical regions depicted in the echo sequences. For example, in apical 4-chamber views, the feature points trace the mitral valve and inner and outer boundaries of the left ventricle. While the feature points are manually isolated during training stage, they are automatically identified during matching.

For each view, a number of training images were collected from cardiac cycle video clips – covering different patients, diseases, and time offset within the cycle. We assume that each echo video clip represents a full heart cycle. Each training image is represented by a shape vector \mathbf{s} obtained by concatenating the locations (x, y) of n features f_1, f_2, \dots, f_n as

$$\mathbf{s} = [x_1, y_1, \dots, x_n, y_n]^T. \quad (1)$$

Next, we form image patches centered around the feature locations to capture texture information. Given a training image with n features, the texture vector \mathbf{t} concatenates the pixels from all the patches into one long vector, where patch size is matched to the pixel spacing between features. Fig. 1 shows the feature points we use for two different cardiac views.

To construct the ASM model, we reduce the dimensionality of the spatial and textural vector using PCA to find a small set of *eigenshapes* and *eigentextures*. The shape \mathbf{s} and texture \mathbf{t} can then be linearly modeled to form the active shape model as:

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

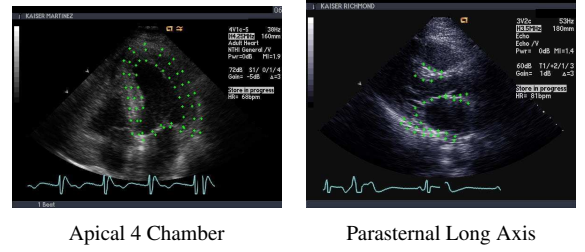


Figure 1. ASM feature points for two cardiac echo views. (Best viewed in color.)

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.

2.2. ASM fitting

Fitting an ASM to a new sequence involves finding a similarity transform Γ^{sim} to position the model appropriately and recovering the shape and texture vectors \mathbf{a} and \mathbf{b} . This is iteratively estimated by alternating between shape and texture update steps. Unlike most ASM techniques, which require a manual initialization to start the model fitting, we use an automatic initialization method where a distance-to-eigenspace method is first used to generate seed ASM initializations.

To evaluate an ASM fit at a given position, we measure error of fit in shape space and texture space using Mahalanobis distance and the reconstruction error suitably normalized. For image I , the ASM fit $(\mathbf{a}, \mathbf{b}, \Gamma^{\text{sim}})$ is

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

where $R = \|\mathbf{t} - T T^T \mathbf{t}\|$, $\mathbf{t} = I(\Gamma^{\text{sim}}(x, y))$, $\lambda_{\text{tex}}^{q+1}$ is the $(q+1)^{\text{th}}$ texture eigenvalue, and Σ_{shp} and Σ_{tex} are diagonal matrices with PCA eigenvalues (see Cootes and Taylor [11]).

2.3. Motion modeling

Using an ASM to track cardiac motion through a cycle with m frames will produce m sets of feature points $\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m\}$, where each \mathbf{s}_i is a column vector of stacked ASM (x, y) feature coordinates. To create a canonical representation, we vectorize the motion in the cardiac cycle by (1) normalizing for image plane geometry by applying the similarity transform from frame 1, Γ_1^{sim} , to all frames, (2) standardizing the time axis to a target length n using piecewise linear interpolation of feature tracks, and (3) factoring out object shape by subtracting out frame 1. This creates our final motion vector \mathbf{m}

$$\mathbf{m} = [\mathbf{s}_2 - \mathbf{s}_1, \mathbf{s}_3 - \mathbf{s}_1, \dots, \mathbf{s}_n - \mathbf{s}_1]^T.$$

If each s_i contains F ASM points, the final vector \mathbf{m} has dimensionality $2F(n-1)$.

Applying PCA to the training set of \mathbf{m} yields a set of eigenmotions \mathbf{e}_i^m and mean motion $\bar{\mathbf{m}}$, giving us a lower dimensional representation

$$\mathbf{m} = M \mathbf{c} + \bar{\mathbf{m}} \quad M = \begin{bmatrix} | & | & & | \\ \mathbf{e}_1^m & \mathbf{e}_2^m & \dots & \mathbf{e}_r^m \\ | & | & & | \end{bmatrix},$$

where r eigenmotions are retained. The r -dimensional weight vector \mathbf{c} is our representation of object motion.

To collect training motions, the manual labeling effort is significant since we must annotate entire sequences. We developed a semiautomatic approach that leverages our ASM tracker. We ran our tracker on a set of training sequences, evaluating the returned tracks and making manual corrections as necessary. After vectorizing the motion vectors and applying PCA, we obtained motion models for our viewpoints of interest.

A new sequence \mathbf{m} can be analyzed to evaluate the degree of fit to our linear motion model. First, we project \mathbf{m} to find \mathbf{c} , computing $\mathbf{c} = M^T [\mathbf{m} - \bar{\mathbf{m}}]$, and then the motion fit is the Mahalanobis distance

$$\text{Motion fit} = \mathbf{c}^T \Sigma_{\text{mot}}^{-1} \mathbf{c},$$

where Σ_{mot} is a diagonal ($r \times r$) matrix of eigenvalues corresponding to $\mathbf{e}_1^m, \mathbf{e}_2^m, \dots, \mathbf{e}_r^m$.

2.4. View recognition algorithm

Putting it all together, Fig. 2 presents our algorithm for recognizing cardiac echo views. We assume that the input sequence I_1, I_2, \dots, I_m represents a single heart cycle, and the motion models were trained assuming that frame I_1 is synchronized with the ECG R peak. For each cardiac view, the algorithm estimates a model fit over the sequence

$$\text{model fit} = \frac{1}{m} \sum_{i=1}^m \text{ASM fit}_i + \text{Motion fit}, \quad (3)$$

which is the average appearance fit over the cycle (eqn (2)) plus the Mahalanobis distance on motion. The recognized view is the one that minimizes this fitting metric.

3. Results

We now describe the results of testing our view recognition algorithm on two data sets.

3.1. Data sets

The first data set came from hospitals in India with physicians recording a complete echo exam in continuous loop videos. Thus no view segmentation or view annotation was present. Starting from video sequences of

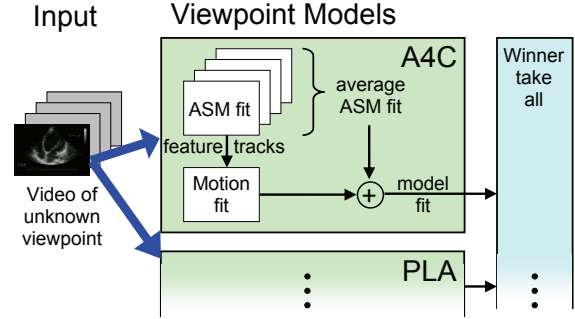


Figure 2. Our algorithm for cardiac echo view recognition.

	India Hospitals		Kaiser	
	# cycles	# frames	# clips	# frames
A4C	26	597	40	1902
PLA	18	384	40	1742
PSAB	12	271	-	-
PSAP	16	360	-	-
Total	72	1612	80	3644

Table 1. We collected two cardiac view data sets for testing, one from a group of hospitals in India, and a second from Kaiser Permanente, Northern California.

the echocardiographer’s entire workflow, we manually extracted individual cardiac cycles for the four views shown in Table 1: parasternal long axis (PLA), apical four chamber (A4C), parasternal short axis - basal level (PSAB), and parasternal short axis - papillary muscle level (PSAP). The videos depicted heart motion of patients covering a range of ages and cardiac diseases. The video was captured at 320x240 and 25 Hz, and a ECG trace waveform at the bottom allowed us to synchronize extracted cycles with the ECG R peak.

The second data set comes from a random draw of cardiology patients from Kaiser Permanente Northern California. The patients span a number of diseases, ages, and Kaiser facilities. Clips were recorded during regular patient cardiology visits, and they typically record 1-2 seconds of video. Frame 1 is often triggered by the ECG R-wave peak, but it is not always synchronized. Clips representing A4C and PLA views were identified manually from DICOM files; 14 patients were used for training ASM models, and a separate set of 40 patients were used for testing (see Table 1).

3.2. Training and testing

Because there is no separate patient training set for the India data, we trained models and performed testing using a leave-one-out methodology. When testing a sequence from patient X, the entire data set minus X is used for training – ASMs and motion models are generated on-the-

	India Hospitals			Kaiser
	ASM+ Motion	ML- Boosting	PMK	ASM+ Motion
A4C	96.2%	80.3%	59.3%	97.5%
PLA	88.9%	75.5%	65.5%	82.5%
PSAB	91.6%	67.5%	47.7%	-
PSAP	75.0%	70.9%	73.2%	-
Ave	87.9%	73.5%	61.4%	90.0%

Table 2. Performance comparison for different view classification algorithms.

fly. For constructing the ASMs, we manually label frames from all sequences, sampling the time axis at regular intervals (about 1/4 cycle period). Motion modeling requires the entire training sequence to be annotated, so we leveraged the ASM model to track the sequences, evaluated and manually corrected the automatic tracks, and used this to bootstrap motion training. After model training, the test patient sequence is fed through view recognition; the results are shown in the first column of Table 2, labeled “ASM+Motion”.

Our Kaiser data set is much larger, so we used separate patient sets for model training and testing. A single ASM and motion model is constructed for A4C and PLA views, and they are used to evaluate the entire test set of 40 patients (2 clips per patient, 1 A4C, 1 PLA). Since the ECG data is sometimes low quality, we cannot assume to have R-wave-based timing for extracting exact cardiac cycles in time. Thus, for the Kaiser data, we drop the motion analysis term from the model fit in Eqn. (3), relying only on the appearance fit. The final column of Table 2 shows the 2-class A4C vs. PLA recognition results.

Finally, for the India data, our ASM+Motion method compares favorably against two state-of-the-art techniques, ML-Boosting [3, 4] and pyramid match kernels (PMK) [12], the results of which are shown in the center two columns of Table 2. PMKs are a generic object categorization technique, which we applied here to the domain of cardiac view recognition.

4. Conclusions

In this paper, we addressed the problem of inferring the transducer-induced viewpoints from echocardiographic sequences. The main differentiating feature of our approach is that feature motion over the cardiac cycle is used for view discrimination. However, if synchronizing ECG data is not available, our system relies on average frame-by-frame ASM appearance fitting. Overall, our spatiotemporal modeling approach allows us to handle the complex echo exam conditions that give rise to large variations in the appearance of heart regions even within known standard views. Comparison with re-implementations of state-

of-the-art view recognition and general category recognition approaches indicate that our technique outperforms these methods while putting minimal constraints on the acquisition of these sequences.

References

- [1] Ebadollahi S, Chang SF, Wu H. Automatic view recognition in echocardiogram videos using parts-based representation. *CVPR 2004* ;2:2–9.
- [2] Otey ME, Bi J, Krishnan S, Rao B, Stoeckel J, Katz A, Han J, Parthasarathy S. Automatic view recognition for cardiac ultrasound images. In *MICCAI-CVII. 2006*; .
- [3] Zhou SK, Park JH, Georgescu B, Comaniciu D, Simopoulos C, Otsuki J. Image-based multiclass boosting and echocardiographic view classification. In *CVPR 2006*; 1559–1565.
- [4] Park J, Zhou S, Simopoulos C, Otsuki J, Comaniciu D. Automatic cardiac view classification of echocardiogram. In *International Conference on Computer Vision. 2007*; 1–8.
- [5] Aschenasy S, Jansen C, Osterwalder R, Linka A, Unser M, Marsch S, Hunziker P. Unsupervised image classification of medical ultrasound data by multiresolution elastic registration. *Ultrasound in Medicine and Biology July 2006*; 32(7):1047–1054.
- [6] Cootes TF, Taylor CJ, Cooper DH, Graham J. Active shape models—their training and application. *Comput Vis Image Underst 1995*;61(1):38–59.
- [7] Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence 2001*;23(6):681–685.
- [8] Beichel R, Bischof H, Leberl F, Sonka M. Robust active appearance models and their application to medical image analysis. *IEEE Transactions on Medical Imaging September 2005*;24(9):1151–1169.
- [9] Zhou SK, Shao J, Georgescu B, Comaniciu D. Pairwise active appearance model and its application to echocardiography tracking. In *MICCAI. 2006*; 736–743.
- [10] 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 Transactions on Medical Imaging November 2002*; 21(11):1374–1383.
- [11] Cootes T, Taylor C. Using grey-level models to improve active shape model search. In *International Conference on Pattern Recognition, volume 1. 1994*; 63–67.
- [12] Grauman K, Darrell T. The pyramid match kernel: Discriminative classification with sets of image features. In *International Conference on Computer Vision, volume 2. 2005*; 1458–1465.

Address for correspondence:

David Beymer
 IBM Almaden Research Center
 650 Harry Rd, San Jose, CA 95129, USA
 beymer@us.ibm.com