ECHOCARDIOGRAM VIDEOS: SUMMARIZATION, TEMPORAL SEGMENTATION AND BROWSING

Shahram Ebadollahi[1], Shih-Fu Chang[1], Henry Wu[2]

[1] Department of Electrical Engineering, Columbia University
[2] The College of Physicians and Surgeons, Columbia University

ABSTRACT

In this paper, we present a system for the temporal segmentation, summarization, and browsing of the Echocardiogram videos. Echocardiogram videos are video sequences produced by the ultrasound scanning of the heart, and are one of the main modalities of imaging the heart structure. Our approach combines the domain-specific knowledge and the automatic analysis of the spatio-temporal structure of the Echocardiogram videos. The videos are temporally sampled using the embedded Electrocardiogram graph. The consecutive sampled frames are compared based on the shape of the Region Of Interest and the presence/absence of color to detect the boundaries between the different segments. The content of each segment of the video is summarized into two forms: the static and the dynamic summaries. Finally the summary is displayed in the user interface in an intuitive form for the purpose of browsing. Applications include digital medical image libraries, medical image management, and tele-medicine.

1. INTRODUCTION

In the current paper, we address the issue of analyzing the spatio-temporal structure of Echocardiogram videos for the purpose of summarization, temporal segmentation, and creation of a digital archive of these videos with efficient and intuitive browsing capabilities. This work also lays the foundation for further exploring the possibility of content-based search in Echocardiogram video libraries.

Echocardiogram video sequences are the product of the ultrasound scanning of the cardiac structure. Due to their low-cost and non-invasive nature, they are a popular imaging modality. Each Echocardiogram video shows the structure of the heart and its dynamics from a sequence of different angles. (The reader is encouraged to see a sample Echocardiogram at: http://www.ctr.columbia.edu/~shahram/demo.html.)

This work is motivated by the growing interest in managing medical image collections based on their content. This interest is justified by 1) the prospect of having the means to search such collections by their content for computer-assisted diagnosis, 2) efficiently browsing such collections, and 3) associating meta-data to the objects of interest in the images/videos.

Currently systems and standards such as PACS [1] and DICOM [2] are being used in the medical imaging centers to digitize, view, communicate, and store medical images. However, these do not take into account the characteristics of the content of the medical images/videos.

In the recent years advances have been made in the content-based retrieval of medical images [3]. Research has also been done on extracting the boundaries of cardiac objects from sequence of echocardiographic images [4]. The image sequences are usually manually selected, and the focus is on finding the exact boundaries of particular objects of interest for the purpose of measuring their characteristics.

The focus of the current work is to look at the Echocardiograms as a video and exploit its specific structure to segment and summarize it. This aspect has not been investigated in detail before.

The analysis of the content of the other types of videos such as films, sports, and news for the purpose of summarizing and indexing is an active field of research [5]. Common approaches at the conceptual level, such as shot segmentation and storyboard summary, can be extended to Echocardiogram videos. However, challenges in developing robust technical methods for Echocardiograms differ from other domains due to the special production characteristics and user task requirements.

In Echocardiograms information is predominantly contained in the visual form. There is usually predictable structure in the production of the Echocardiogram videos. In addition there is associated information from other modalities, such as the diagnosis reports, and ECG (Electrocardiogram shows the electrical activity of the heart. In Echocardiograms, the ECG is embedded as a moving graph in the frames of the video.), which can be used in analyzing the video content or providing useful annotations.

The current paper reports the following unique contributions:

- A framework for the automatic detection of the boundaries between the views of an Echocardiogram video.
- Sampling the content of the Echocardiograms using the embedded ECG visual graph and defining video summaries based on these samples.
- An efficient and intuitive interface for browsing the content of the videos with a 3D graphic model of the heart.
- Use of the domain-specific knowledge to analyze the spatio-temporal structure of the Echocardiograms.

The techniques described here are being used in constructing the Columbia's Digital Echocardiogram Video Library (DEVL), which will be discussed later in this paper.

In section 2, we will introduce the structure of the content of the Echocardiograms, and will also define some domain-specific terminology. The system for segmenting and summarizing the Echocardiograms is presented in section 3. Section 4 will be about the user-interface and browsing of the content of the Echocardiograms. Results and conclusions will be summarized in section 5.

2. ECHOCARDIOGRAM VIDEOS: DEFINITION AND CONTENT STRUCUTRE

An Echocardiogram is a video sequence obtained by scanning the cardiac structure with an ultrasound device [6]. The scanning device is positioned in certain locations on the patient's chest. Thus the Echocardiogram shows the cardiac structure from a sequence of different angles corresponding to the successive transducer locations. The locations, their total number, and their order are more or less fixed, and follow the recommendation by the *American College of Cardiology*. Variations may exist among different imaging centers in acquiring the Echocardiograms. The duration of the video segment corresponding to each angle/location is not fixed. The total duration of an Echocardiogram is 10 to 15 minutes.

At every angle, different modes of echocardiography are possible. Transition from one mode to the other is probabilistic and depends on the case. These modes are:

- *2-Dimensional (2D)*, which is the gray-scale reference frame sequence at each view angle.
- *Color Doppler*, which uses color overlays on the reference frame sequence to show the blood flow in the heart based on the Doppler effect.
- *Zoom-in*, which is a closer view of the Region of Interest (ROI).
- *Pulsed-Doppler*, which shows the blood flow in a particular direction.

We define a *view* to be a sequence of frames corresponding to a single transducer location and mode of imaging. A view can be regarded as the equivalent of a *shot* in general videos. The transition from one view to another is always abrupt, though with subtle differences between the frames at the either side of the transition, with or without blank frames between them. The sequence of views under each view angle always starts with the 2D view, followed by other optional modes including the 2D itself. The optional modes may repeat in random order.

The spatial structure of the frames in a particular view is determined by the mode of echocardiography in that view. Figure 1 shows this spatial structure for two of the modes. Each frame consists of a ROI, text overlay, dynamic range indicator, and an ECG graph, except for the Pulsed-Doppler views, which do not have an ECG. In our system we exploit the shape of the ROI, presence/absence of color, and the information extracted from the ECG to segment and summarize the videos.

Finally it is worth looking at the ECG in more detail because we use it to sample the content of the Echocardiogram, i.e. to extract key-frames for each view of the video. The ECG graph embedded in the frames of the video as seen in figure 2 is not useful for diagnosis, but is important as a timing reference. Usually for diagnosis purposes a set of 12 ECG signals are used which are not available here.

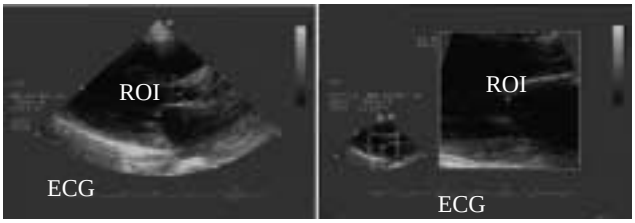

Figure 1. The image on the left shows a standard 2D frame and the one on the right is a frame from a zoom-in mode in the same view angle.

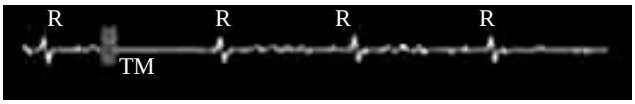

Figure 2. Image of the ECG as it appears at the bottom of the frames. The Time-Marker and the R-Peaks are marked with *TM* and *R* respectively. The Time-Marker sweeps the graph from left to right as time progresses.

Each view is consisted of several heart cycles. Each cycle consists of two phases. One is *Systole* where heart is contracting, and the other is *Diastole* where heart is expanding. The ECG signal can be used to track these phases. One important time instance in our analysis is the point where heart is most expanded (*End-Diastole)*. This is a frame in each heart cycle where the *Time*-Marker on the ECG graph has just passed over a peak (*R-peak*). See figure 2 for illustrations.

3. SEGMENTATION AND SUMMARIZATION

In this section we use the domain-specific knowledge about the Echocardiogram video and its unique spatio-temporal structure to sample the content of the videos, automatically detect the boundaries between the views and summarize the contents of each view. The architecture of the proposed system is shown in figure 3.

3.1. Content-Based Sampling of the Echocardiogram Video

Content-based sampling of the video has been used to obtain a subset of video frames; the so-called *Key-Frames*, that contains the salient information of the video. Here we sample the Echocardiogram videos for two reasons:

First, in the process of view-boundary detection instead of comparing features of the consecutive frames of the video, we only compare the sampled frames, which helps in speeding up the operations. Second, in order to summarize the content of each view in the Echocardiogram, we need to extract certain frames of the view that best capture its visual information.

At the End-diastole of the heart cycle, the heart is most expanded. In this state different objects in a particular view are viewable with good detail. Therefore the *End-Diastole* frames provide an appropriate visual reference to the content of the view. We use the available *ECG* graph to sample the Echocardiogram video at the time instances corresponding to the *End-Diastole*.

The time-marker (*TM*) is continuously sweeping the *ECG* from left to right. By tracking the position of the *TM* in the sequence of frames and comparing this location to the positions of the *R-Peaks* we can find the time instances when the *TM* passes over a *R-Peak*.

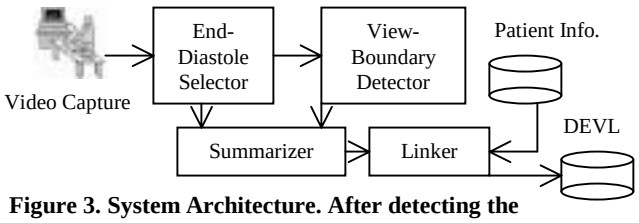

Figure 3. System Architecture. After detecting the Key-frames of the video, those key-frames are analyzed to detect possible view boundaries. Using the key-frames of each view, and the frames corresponding to one cycle of the heart video is summarized. The linker associates the related diagnosis reports to the summary of each view of the video.

First, we extract the locations of the *TM*s. (Note: sometimes there are multiple *TM*s on the *ECG*. In this case only one of the TMs is moving and the others are static. The dynamic TM is the important one.) For this, we do a morphological opening on the image of the *ECG* area with a square structuring element. Then the result is closed morphologically by an up-right rectangular structuring element. This operation will identify the possible locations of the *TM*s on the graph. The possible *TM*s are filtered based on their area. The ones that don't fall between certain thresholds are rejected. The centroids of the *TM*s are calculated next and by comparing the centroids of the TMs in the current frame and the previous frame the *TM*s are classified into static or dynamic. We only need to keep track of the dynamic ones for the purpose of locating the End-Diastole frames in each view.

The identified *TM*s are subtracted from the image of the *ECG* graph before trying to find the location of the R-peaks. Morphological dilation with a vertical line is applied to the result image in order to emphasize the locations of the *R-Peaks*. Based on the ECG embedded in the Echocardiogram video, the peaks of the *ECG* in the video can be either pointing up-ward or down-ward. In order to take this into account we morphologically erode the image with two structuring elements, one is an arrow pointing upward and the other a downward one. The result of this operation is a cluster of pixels around the location of each *R-Peak*. We need to distinguish between these clusters of points, and replace each of them with their centroids to represent the R-peaks. K-means clustering algorithm is used for this. The criteria for stopping the clustering algorithm is the ratio of the inter to intra cluster distances.

By counting the number of *R-Peaks* to the left of the dynamic *TM* and keeping track of this count with time, we can declare a frame to be the *End-Diastole* when the count increases by one. In order to avoid false positives we filter the graph of the count versus frame number using a heuristic method.

We detect more than 95% of the *End-Diastole* frames using this method in the majority of the videos. Sometimes due to acquisition problems the *ECG* graph does not have good quality which causes false positives.

3.2. View Boundary Detection

Various algorithms have been proposed in the literature for shot boundary detection in general videos, such as frame difference, histogram difference, edge correlation, … Due to high speckle noise present in the Echocardiogram videos it is difficult to apply these algorithms to the Echocardiograms. We use the unique features of the different views in the Echocardiogram to detect the view boundaries.

In parallel to the content-sampling process, the extracted features of the sampled frames are continuously being compared with each other to detect the view boundaries. These features are the shape of the *ROI* and the presence/absence of color in the sampled frames. The reader is referred to [7] for more details on extracting these features in the sampled frames.

Based on the combination of the shape and color features the mode of the frame is determined. The view boundaries correspond to the discontinuities in the mode of the sampled frame sequence. The detected view boundaries have an error of less than the duration of a heart cycle, which is sufficient for clinical purposes (the exact location can also be found by comparing the consecutive frames, but the result is not justified by the additional cost in the operations).

By using this method we can detect 90% of the view boundaries on average (the exact measure varies in different Echocardiograms). The missed 10% are due to the times, when 2D views are placed one after the other without any other modes between them or any blank frames separating those views. The way to distinguish between these consecutive 2D views is to recognize the view angle they belong to. This is possible by finding the salient features of these 2D views.

We are currently working on an algorithm which uses the number of the cavities, their locations, and the way these change throughout one heart cycle to identify the different 2D views, i.e. to determine which view angle they correspond to. We use the gray-level symmetric axis transform which is a method for describing objects in different resolution levels to locate the cavities in each frame of a heart cycle. Each cavity corresponds to a chamber in the heart. In different views different combinations of chambers are viewable. Note that we only need to determine the approximate locations of the chambers of the heart in this process. See figure 4 for an illustration of this idea.

3.3. Echocardiogram Summarization

Video summaries are helpful in browsing the content of the digital video libraries. One type of the summary is the *static summary*, which usually is a set of representative frames of the video presented to the user. This type of the summary gives a general overview of the content of the video and provides random-access entry points to the content of the video. There are also video skims and video abstracts, which present an informative shorter version of the video to the user.

For the Echocardiogram videos we define two types of summaries. The first one, which is called the *Static Summary*, is a collection of the representative End-Diastole frame of each view for all the views of the video. This type of summary is useful for randomly accessing the content of the Echocardiogram and for linking the views to the textual reports, which are associated with the video.

The second type of the summary, the *Dynamic Summary*, is a collection of short segments extracted for different views of the video. The extracted segment is a sequence of frames that correspond to one heart cycle. A cycle is a sequence of frames between two consecutive R-Peaks. Since there are multiple cycles per view, the last one is chosen to represent the view in the dynamic summary. This is because usually at the end of each view the location of the transducer is stabilized and the

extracted segment could be regarded as reliable in terms of representing the correct content.

This type of summary is also called "*Clinical Summary*" because the video of one cycle is clinically sufficient for conveying the information about that view. This type of summary is useful in tele-medicine applications.

4. BROWSING THE CONTENT OF THE ECHOCARDIOGRAM VIDEO

We have used the system mentioned here to build a Digital Echocardiogram Video Library (DEVL) of sample cases. There are approximately 50 different cardiac diseases under five main categories that one can diagnose using the Echocardiograms. Our library consists of 750 Echocardiogram videos; 15 under each category of disease.

In the interface of the library the user can browse the disease categories to locate the disease of his/her interest. The disease categories are arranged in the form of a tree structure with the main categories at the very top. After finding a disease, the user can view an Echocardiogram under that disease category by selecting one of the patient's videos. This will take the user to an interface for browsing the content of the Echocardiogram (Figure 5).

Each button on the left-hand side of the screen represents a different view angle of the video. On the right-hand side of the screen there is a 3D model of the heart that the user can drag and rotate. By choosing a particular view angle from the left-hand buttons, the representative frames of different modes of echocardiogram associated with that view angle will be displayed. At the same time a cutting plane cuts the 3D model of the heart to show the approximate view angle of the transducer. The user can click on any of the frames to see the particular view that the frame is representing in the summary.

Such an interface for browsing the content of the Echocardiograms is intended for expert users such as the doctors. It is also useful in training the medical students. In this sense it can be viewed as a digital atlas of sample disease cases, where the medical student can browse and look at the visual features of each case.

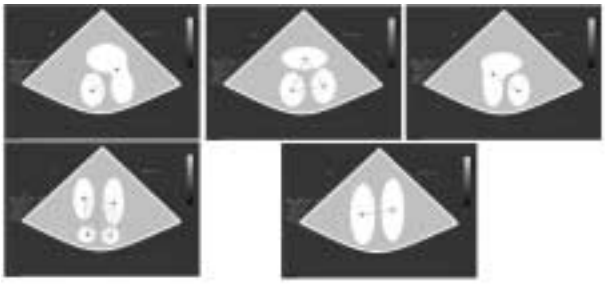

Figure 4. Each row shows the different combinations of cardiac chambers as heart goes through one cycle from left to right. Top row corresponds to "Apical 3 Chamber" view, and the bottom row to "Apical 4 Chamber" view.

6. CONCLUSIONS

In this paper, we presented a systematic and effective framework for structure analysis and summarization of the Echocardiogram videos, which are one of the most popular modalities of cardiac imaging. Our approach uses the unique spatio-temporal structure of the videos and the embedded ECG graph for view boundary detection, representative frame extraction and summarization. An application of the system in building a digital video library of the Echocardiograms was also presented. Future work is automatic recognition of the different views in the Echocardiogram video, annotating the cardiac objects, and content-based retrieval for this type of content. The applications of this work are in digital medical image libraries, medical image management, and tele-medicine. We haven't extensively evaluated the system presented in this paper in actual clinical environment. We are collaborating with echocardiography labs to do the evaluations.

6. REFERENCES

[1] Huang, H.K., *PACS: Basic Principles and Applications*, Wiley, New York, 1999.

[2] DICOM: http://medical.nema.org/dicom.html

[3] C.R. Shyu, C.E. Brodely, A.C. Kak, and A. Kosaka, "ASSERT: A Physician-in-the-Loop Content-Based Retrieval System for HRCT Image Databases," *Computer Vision and Image Understanding*, Vol. 75, Nos. ½, July/August, pp. 111-132, 1999.

[4] J.S. Duncan and N. Ayache "Medical Image Analysis: Porgress over Two Decades and the Challenges Ahead," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 1, January 2000, pp.85-105, 2000.

[5] S.F. Chang and H. Sundaram "Structural and Semantic Analysis of Video," *ICME 2000, New York, New York*, July 28 – August 2, pp.687-690, 2000.

[6] Feigenbaum, H., *Echocardiography*, LEA & FEBIGER, 1993.

[7] S. Ebadollahi, S.F.Chang, H. Wu, S. Takoma "Echocardiogram Video Summarization," Proceedings. *SPIE Medical Imaging 2001: Ultrasonic Imaging, Vol. 4325, San Diego, CA.*, February 2001, pp.492-500, 2001.

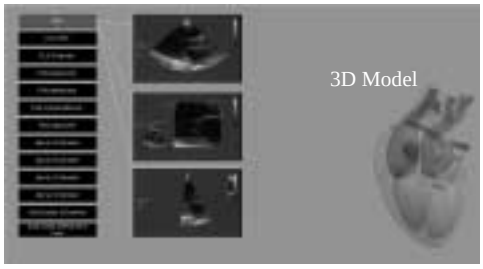

Figure 5.Browsing interface. On the left there is a button for each view angle in the video. By clicking on one of them the corresponding representative frames are displayed in the center. A 3D plane cuts the heart model on the right side to show the corresponding cross section of the heart.