Medical Image Analysis 36 (2017) 15–21

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Automatic apical view classification of echocardiograms using a discriminative learning dictionary

Hanan Khamis[a,1,*], Grigoriy Zurakhov[a,1], Vered Azar[a], Adi Raz[a], Zvi Friedman[b], Dan Adam[a]

[a] Department of Biomedical Engineering, Technion - IIT, Haifa, Israel
[b] GE Healthcare, Tirat Hacarmel, Israel

ARTICLE INFO

Article history:
Received 20 April 2016
Revised 14 October 2016
Accepted 22 October 2016
Available online 24 October 2016

Keywords:
Echocardiography
Echocardiogram classification
Cuboid-detector
Supervised dictionary learning
LC-KSVD

ABSTRACT

As part of striving towards fully automatic cardiac functional assessment of echocardiograms, automatic classification of their standard views is essential as a pre-processing stage. The similarity among three of the routinely acquired longitudinal scans: apical two-chamber (A2C), apical four-chamber (A4C) and apical long-axis (ALX), and the noise commonly inherent to these scans - make the classification a challenge. Here we introduce a multi-stage classification algorithm that employs spatio-temporal feature extraction (Cuboid Detector) and supervised dictionary learning (LC-KSVD) approaches to uniquely enhance the automatic recognition and classification accuracy of echocardiograms. The algorithm incorporates both discrimination and labelling information to allow a discriminative and sparse representation of each view. The advantage of the spatio-temporal feature extraction as compared to spatial processing is then validated.

A set of 309 clinical clips (103 for each view), were labeled by 2 experts. A subset of 70 clips of each class was used as a training set and the rest as a test set. The recognition accuracies achieved were: 97%, 91% and 97% of A2C, A4C and ALX respectively, with average recognition rate of 95%. Thus, automatic classification of echocardiogram views seems promising, despite the inter-view similarity between the classes and intra-view variability among clips belonging to the same class.

1. Introduction

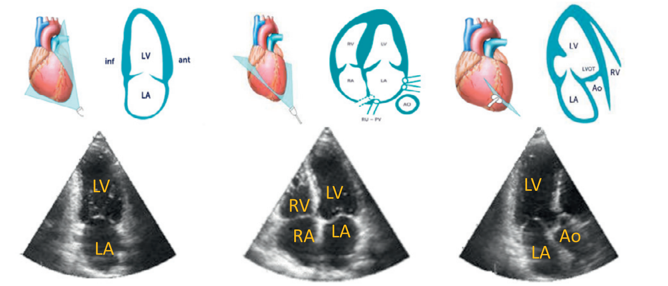
Current echocardiographic software packages (e.g. EchoPAC (GE healthcare), QLAB (Philips), etc.) for cardiac functional analysis require various processing algorithms in order to provide a full and reliable assessment of the cardiac functionality. These software packages may involve algorithms for segmentation, detection of anatomical biomarkers, blood/tissue tracking, etc. In addition, in the clinical practice, images from multiple modalities are managed and stored in the widely used Picture Archiving and Communication Systems (PACS). Clinicians manually choose the required image for analysis and diagnosis. Despite the efforts that have been invested in the automation of these algorithms, they usually require user interaction, and frequently necessitate human involvement in recognition of the echocardiogram views. Since the echocardiogram views are noisy, and share similar shape information, it might be challenging and exhausting to classify large databases and correctly identify the views, which may lead to unreliable or incorrect analysis. For example, 2D speckle tracking echocardiography algorithms, by Leitman et al. (2004), require a prior information regarding the analyzed view.

Hence, a fully automatic and reliable classification of echocardiogram views is considered as a mandatory initial step to subsequent automatic analysis of the clips, and as well as a quality check tool. Furthermore, automatic apical view classification of echocardiograms may be very useful for pre-labeling large databases of unclassified images, or as part of a fully automated analysis chain. This may be a useful tool e.g. for better control of classification errors due to human factors (Rigling, 2007), or for advancing automatic echocardiographic point of care applications both in the field and at the bedside.

Standard echocardiogram views acquired during a routine clinical echo exams (as recommended by the Guidelines (Lang et al., 2015)) are views scanned through the apical and parasternal acoustical windows. There are 4 apical views: apical two chamber (A2C), apical four chamber (A4C), apical long-axis (ALX), and apical five chamber (A5C) views. Additionally, There are 2 main parasternal views: long-axis (PLAX) and short-axis (SAX) views, where the short axis views can be acquired at 3 main levels: mitral valve (MV), papillary muscle (PM) and apex (AP) views (Lang et al., 2015). In this work, the focus is on three apical views: A2C, A4C and ALX (Fig. 1).

* Corresponding author.
E-mail address: khamishanan@gmail.com (H. Khamis).
[1] Equal contribution.

http://dx.doi.org/10.1016/j.media.2016.10.007
1361-8415/



(a) Apical 2 Chamber view (A2C) (b) Apical 4 Chamber view (A4C) (c) Apical Long Axis view (ALX)

Fig. 1. Echocardiographic apical views: (a) Apical 2 Chamber view (A2C), (b) Apical 4 Chamber view (A4C) and (c) Apical Long Axis view (ALX). (Courtesy and copyrights: 123sonography.com)

Echocardiographic clips and images are characterized by several properties that make the classification task a challenge. Among them are:

I. The intra-view variability of echocardiograms of the same cardiac view, due to physiological variations among subjects, different acquisition parameters (angle, depth, properties of the scanning machine, foreshortening, etc.) and the sonographer's expertise.
II. The inter-view similarity of echocardiograms of different cardiac views, due to similar information in both views (such as valve motion, wall motion, left ventricle, etc.), in addition to ill-defined transducer position during the acquisition, that may lead to imprecise capture and ambiguous view.
III. The redundant information that appears in all echocardiograms independent of the view, such as exam information (date and time of exam, ECG, heart rate, frame rate) and the scanner details, which may corrupt the classification process.

In addition, speckle noise and clutter noise lower the clarity of the images thus limiting the ability to perform accurate view classification. Dropout phenomena also makes the classification challenging.

Object recognition and classification, in general, are well-known challenges in the field of computer vision. Several efforts have been made to achieve accurate recognition; among them are dictionary-learning and machine-learning based algorithms, which have been shown to outperform other methods (Jiang et al., 2013). There is, though, a limited number of publications in the literature that are directly related to classification of echocardiogram views. For example, Ebadollahi et al. (2004) suggested to sub-divide the heart into its chambers, by using part-based representation approach. The spatial and statistical properties of the chambers are then modeled by Markov Random Fields. These models are used to represent echocardiograms and to classify them into categories using a support vector machine (SVM) algorithm. This technique may fail when applied to images in which extra/less chambers seem to appear due to high level noise or different acquisition depths. Ten different views belonging to parasternal views (PLAX and PSAX) and apical views were used, for normal and abnormal echocardiograms, but no specific classes were mentioned. The reported classification accuracy of this technique was of 88.35% for the normal views and 74.34% for the abnormal views. A different approach was reported by Otey et al. (2006) who utilized, for feature extraction, the magnitude of the gradient in space-time domain of the echocardiogram clips, followed by a hierarchical classification scheme: first classification into apical views and PSAX, then classification into the sub-views. Here, the ALX view was not included in the classification process. The total classification accuracy of A2C and A4C views was 88.7% for the leave-one-out cross validation and 100% for the testing data (the latter is composed of only 14 clips of each view). Aschkenasy et al. (2006) suggested a landmark-free and unsupervised classification of echocardiogram clips by using a multi-scale elastic registration algorithm. A 3rd order direct B-spline transform filter was used to reconstruct multi-scale template images, representing the different views. The total classification accuracy of A4C and A2C views was 85.7%. One major limitation of this technique is its dependency on the templates chosen specifically for each view, which might be sensitive to the variability between operators, scanners and subjects. In another work, a supervised machine learning approach was used by Park et al. (2007). They train a detector for each view of the left ventricle, using Haar wavelet type local features and a 'multi-class boosting' learning technique. The total classification accuracy of this technique for the A4C and A2C was 95.7%. These aforementioned studies by Otey et al. (2006), Aschkenasy et al. (2006) and Park et al. (2007) focused on classification between the views (A2C, A4C, PLAX, PSAX).

Agarwal et al. (2013) used histogram of oriented gradients as the discrimination features for encoding the spatial arrangement of edges/gradients in the images. This information was later used as an input to the SVM classifier. This study, though, focused only on classifying between PLAX and PSAX views. A different approach was suggested by Qian et al. (2013), in which they used "bags of words" coupled with linear SVM's. They used sparse coding method to train an echocardiogram video dictionary, based on 3D SIFT descriptors of space-time interest points, which were detected by a Cuboid detector. The linear multiclass SVM was used to classify echocardiogram clips into eight views. In this study the following views [A2C, A4C, ALX, A5C, PLAX, PSAX view of the Aorta, PM and MV] were included, where the average classification accuracy of A2C, A4C and ALX was 68%. One may notice that 79% of the classification errors were within the apical views category.

It should be noted that just a few studies have attempted to classify concurrently the three apical long-axis views, which should be studied according to the guidelines (Lang et al., 2015). Classification into the three standard apical views is a challenging task due to the inter-view similarity, intra-view variability and presence of noise (stationary and dynamic clutter, decorrelation noise, etc.). Nevertheless, it is still a highly important task required by the clinicians.

Recent developments of new algorithms in the field of machine and dictionary learning, and the development of advanced computer vision techniques allow concurrent enhancement of the image classification accuracies. Thus, here we propose a multi-stage process of classification, by first using the cuboid detector for spatio-temporal feature extraction (Section 2.2.2) followed by an employment of the Label Consistent K-SVD algorithm (LC-KSVD), proposed by Jiang et al. (2013) (Section 2.2.3), to represent the features while forcing discriminative sparse coding to enable better recognition accuracies. The LC-KSVD algorithm was selected here since it was reported (Jiang et al., 2013) to outperform many sparse-coding based techniques for the recognition of face, action, scene, and object categories.

The difficulties encountered when attempting to classify the echocardiographic views, motivated us to first search for visual cues, by studying a large set of echocardiograms. Both spatial and temporal information, such as location of anatomical markers and their temporal motion may serve as visual cues (features), which may lead to better classification accuracies. Prior visual study of the three echocardiographic apical views has taught us that the main distinguishable morphological differences between the views are usually located almost at the same depth as the mitral valve (MV). The aortic valve (AV) and aorta can be visually detected only in the ALX view, while the right chambers and the tricuspid valve (TV) can be visually detected only in the A4C view. Hence, we propose to use only the MV region in the classification process, where