

Research Article

Dense Trajectories and DHOG for Classification of Viewpoints from Echocardiogram Videos

Liqin Huang, Xiangyu Zhang, and Wei Li

School of Physics and Information Engineering, Fuzhou University, Fuzhou 350116, China

Correspondence should be addressed to Liqin Huang; lquangfzu@163.com

Received 15 October 2015; Revised 19 January 2016; Accepted 31 January 2016

Academic Editor: Syoji Kobashi

Copyright © 2016 Liqin Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In echo-cardiac clinical computer-aided diagnosis, an important step is to automatically classify echocardiography videos from different angles and different regions. We propose a kind of echocardiography video classification algorithm based on the dense trajectory and difference histograms of oriented gradients (DHOG). First, we use the dense grid method to describe feature characteristics in each frame of echocardiography sequence and then track these feature points by applying the dense optical flow. In order to overcome the influence of the rapid and irregular movement of echocardiography videos and get more robust tracking results, we also design a trajectory description algorithm which uses the derivative of the optical flow to obtain the motion trajectory information and associates the different characteristics (e.g., the trajectory shape, DHOG, HOF, and MBH) with embedded structural information of the spatiotemporal pyramid. To avoid “dimension disaster,” we apply Fisher’s vector to reduce the dimension of feature description followed by the SVM linear classifier to improve the final classification result. The average accuracy of echocardiography video classification is 77.12% for all eight viewpoints and 100% for three primary viewpoints.

1. Introduction

The echocardiography video still plays an important role in modern medical diagnosis. It can be used to analyze the heart by providing the cardiac structural and motional information. The echocardiography video gets the 3D detailed anatomical structure and functional information of the heart from eight standard views, which are usually taken from an ultrasound transducer at the three primary positions (Apical Angles (AA), Parasternal Long Axis (PLA), and Parasternal Short Axis (PSA)). In order to choose the most helpful echocardiography video for disease diagnosis or research, we need to categorize the eight kinds of echocardiography videos. Classifying is the first step in the research. In clinical practice, because of the ultrasound characteristic limits, such as the high noise caused by the low contrast ratio in echocardiography video, sonographer has to manually classify these echocardiography videos. It causes the great decrease of working efficiency and easily impacts the recognition results owing to sonographer experience and image resolution. Therefore, how to use the computer to classify the

echocardiography video is a crucial step in the current echocardiographic research.

In recent years, there are a lot of researchers who focus on this area. For example, Kumar et al. [1] are devoted to expressing the main anatomical structure of space motion. Zhou et al. [2] and Beymer et al. [3] mention echocardiogram classification based on multiple object space relations. Then Shalbfaf and Sera et al. [4, 5] add the movement information on this basis, which extract and detect the feature by tracking the movements and patterns of the contour shape of the heart. Wang et al. [6] extract feature by analyzing the factor in video. Guo et al. [7] propose an echocardiography video classification method based on sparse representation and greatly enhance the recognition accuracy. An echocardiography video classification method based on the 3DSIFT feature description is proposed in [8]. The authors detect echocardiographic features applying cuboid detector, represent these features into 3DSIFT descriptors, and finally use Bag of Word to encode before classifying in SVM. After that, Rap and Espinola-Zavaleta et al. [9, 10] propose the echocardiography classification method based on the features of low dimension

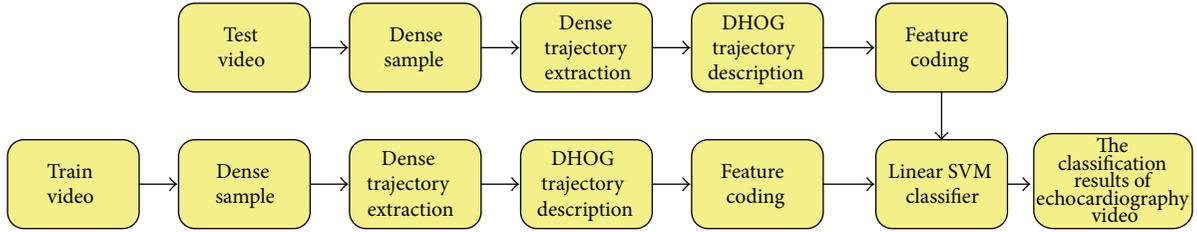


FIGURE 1: Echocardiography video classification flow based on the dense trajectory.

and improve the classification result by combining the kernel function with movement decomposition algorithm.

In this paper, it is a challenge to precisely classify the echocardiography video because it has strong noise and low resolution. We propose a method based on the dense trajectory [11] and difference histograms of oriented gradients (DHOG). Dense trajectory tracking, which samples the video points densely in our experiment, can ensure the feature points covered integrally. In addition, it can differentiate the foreground and the background from the scene and improve the efficiency of recognition. Moreover, dense optical flow can improve perfectly the trajectory characteristics of the feature points. In order to describe the dense trajectory, this paper also proposes a method based on motion boundaries and structure descriptors (DHOG, HOF, and MBH), which has the better result. The movement boundary histogram has higher robustness than other descriptions which are based on the optical flow algorithm. It has a great impact for the stability of the classification system. Finally the Fisher vector and linear SVM classifier are applied to classify echocardiography videos.

2. Methodology

Dense trajectories are obtained through dense-sampling the points in each frame and track these points in dense optical flow field. Dense sampling ensures the integrity coverage of the feature points of video, and dense optical flow can improve the properties of trajectory, distinguish the foreground and background of echocardiography video, and promote the efficiency of recognition. The flow chart of dense trajectory behavior recognition method is shown in Figure 1. We will produce details about the flow chart in the following content.

2.1. Dense Sampling. Feature points on a grid spaced are densely sampled by W pixels. Sampling is carried out on each spatial scale separately. It means that feature points equally cover all spatial positions and scales. According to the previous experiment results [7], the sampling step size of $W = 5$ pixels is dense enough to give good results over all data. Smaller sampling step does not have more improvement and it will increase the amount of calculation. There are at most 8 spatial scales in total, depending on the resolution of the video. The spatial scale increases by a factor of $1/\sqrt{2}$.

Because sampling in the homogeneous area makes no sense for detecting, we remove points in these areas using

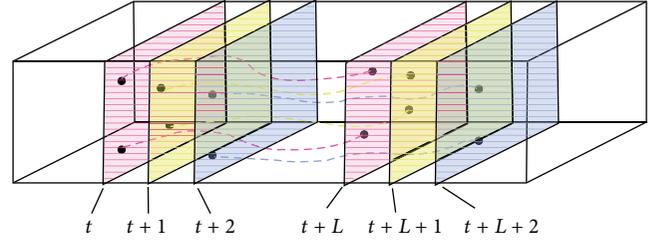


FIGURE 2: The diagram of dense trajectory extraction.

Zhang's criterion [19]. When the eigenvalue of the local autocorrelation matrix is less than a certain threshold, the point will be removed according to the criterion. In this paper, the threshold T is as follows:

$$T = 0.001 \times \max_{i \in I_t} \min(\lambda_i^1, \lambda_i^2), \quad (1)$$

where $(\lambda_i^1, \lambda_i^2)$ are the eigenvalues of point i in the image I_t . And experimental results showed that the value of 0.001 represents a good compromise between saliency and density of the sampled points.

2.2. Dense Trajectory Extraction. Dense trajectories are extracted by tracking feature points in each scale space alone, as shown in Figure 2.

For each frame I_t , its dense optical flow field $\omega_t = (u_t, v_t)$ is computed by w, r, t . The next frame is where u_t and v_t are the horizontal and vertical components of the optical flow. Given a point $P_t = (x_t, y_t)$ in frame I_t , its tracked position in frame I_{t+1} is smoothed by applying a median filter:

$$P_{t+1} = (x_{t+1}, y_{t+1}) = (x_t, y_t) + (M * \omega_t)|_{(x_t, y_t)}, \quad (2)$$

where M is the median filtering kernel and the size is 3×3 pixels. It avoids trajectories of points located on motion boundaries being smoothed out [20].

Once the dense optical flow field is computed, points can be tracked very densely without additional cost. Another advantage using the dense optical flow field is the smoothness constraints which allow relatively robust tracking of fast and irregular motion patterns. In this paper, we use Li's algorithm [21] to embed a translation motion model between neighborhoods of two consecutive frames. And polynomial expansion is employed to approximate pixel intensities in the neighborhood. Not only can this method effectively

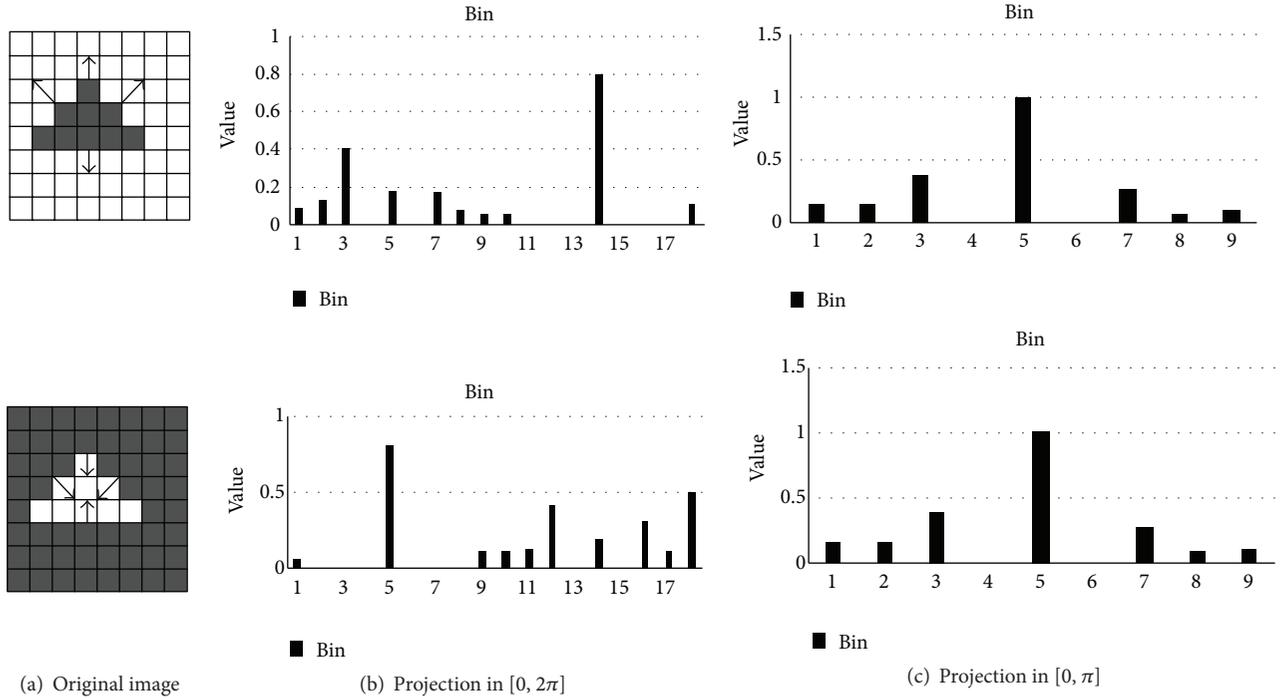


FIGURE 3: The robustness of the compression direction on background.

cooperate with optical flow and improve the accuracy of the optical flow, but also its principle is simple and it has low computational complexity. In our research, we use the implementation from the OpenCV library to complete the calculation of dense optical flow field.

The shape descriptor of a trajectory means encoding local motion positions. Given a trajectory of length L , we describe its shape by a sequence $(\Delta P_t, \dots, \Delta P_{t+L-1})$ of displacement vector $\Delta P_t = (P_{t+1} - P_t) = (x_{t+1} - x_t, y_{t+1} - y_t)$. The resulting vector is normalized by the sum of displacement vector magnitudes:

$$T = \frac{(\Delta P_t, \dots, \Delta P_{t+L-1})}{\sum_{j=t}^{t+L-1} \|\Delta P_j\|}. \quad (3)$$

In the following, we refer to this vector as trajectory. As we use trajectories with a fixed length of $L = 15$ frames, we obtain a 30-dimensional descriptor.

2.3. Motion and Structure Descriptors. Besides the trajectory shape information, we also design descriptors to embed appearance and motion information and include difference histograms of oriented gradients (DHOG), histograms of optical flow (HOF), and the motion boundary histograms (MBH). These features are combined as a motion model of the local feature descriptor. In L frames of continuous video, we extract the feature in the $N \times N$ pixels of local area. In order to embed structural information, we divide the $N \times N \times L$ spatiotemporal cube into a set of grids with $n_\sigma \times n_\sigma \times n_\tau$ and compute the local descriptor corresponding to each grid. Finally, we combine each descriptor into the final vector.

2.3.1. Different Histograms of Oriented Gradients. HOG feature [22] is the information description based on the local statistical characteristics; its main thought is that the local target appearance and shape in an image can be described by the gradient or the direction of the edge density distribution. First the image is divided into several nonoverlapping cells with size $N \times N$, and the gradient of each cell is calculated and then represented using histogram statistics. By merging multiple cells into a vector, the HOG features are obtained.

Although the traditional HOG features have strong robustness for certain amount of illumination change, the expression ability of HOG feature still exists. Since the gradient direction can be changed by the light and shade of the background, it is possible to generate different expressions for the HOG feature as shown in Figure 3. The projection in the interval $[0, \pi]$ has no change for the gradient direction histogram, as shown in Figure 3(c).

It can lead to ignoring the difference information of some target showed. The HOG feature descriptor is uniform because of the different shape objects with different opposite direction of gradient and the same size and number, which can produce the same histogram. So the HOG features have no distinguished ability for these objects.

In order to improve the expression ability of HOG feature in echocardiography video characteristic trajectory. We propose an improved HOG descriptor algorithm. We divide L into $[0, 2\pi]$ area and get the gradient direction histogram H_g , where L is even and each pixel in the cell needs to be vote:

$$H_{og}(i) = H_g(i) + H_g\left(i + \frac{L}{2}\right), \quad 1 \leq i \leq \frac{L}{2}, \quad (4)$$

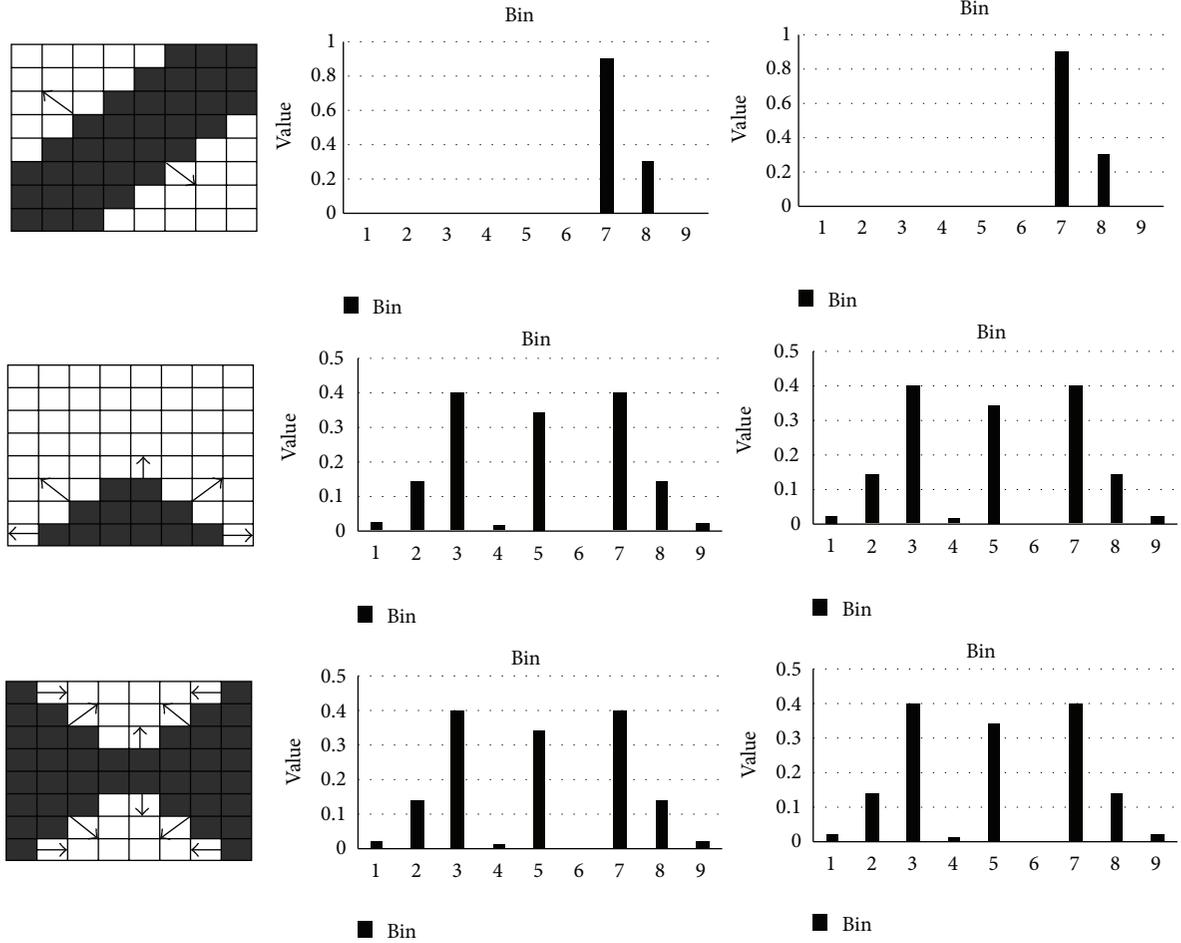


FIGURE 4: Comparison of HOG and DHOG.

where $H_{og}(i)$ and $H_g(i)$ are, respectively, i th element values of HOG and H_g . In this paper, we set a new series of histograms H_{ng} whose length is equal to the HOG after the HOG, as shown in Figure 4, and each element value is

$$H_{ng}(i) = \left| H_g(i) - H_g\left(i + \frac{L}{2}\right) \right|, \quad 1 \leq i \leq \frac{L}{2}. \quad (5)$$

In this paper, we call the new gradient direction histogram as difference HOG (DHOG); in order to simplify the calculation, we extract the HOG feature descriptors in 32×32 pixels of dense trajectory. Each block is divided into 2×2 cells. In order to obtain the optimal description while considering the differences in the opposite direction in this paper, the area ($0 \sim 180^\circ$) is divided into nine bins and then extended to $[0, 2\pi]$ by DHOG. We divide the gradient direction area of $0 \sim 360^\circ$ into 18 bins, and calculate gradient direction histogram in each cell. Then we divide the $L = 15$ frames' dense trajectory into three parts and sum the DHOG features with the corresponding bin. The eigenvectors also are normalized with the L_2 -norm. So, in this paper, there is a dense trajectory of HOG feature with $L = 15$ frames, whose dimensionality is $2 \times 2 \times 18 \times 3 = 216$.

2.3.2. Gradient and Optical Flow Histograms. Different from the DHOG descriptor extracting the static shape information, the HOF uses optical flow to describe and extract the motional information. The calculation method of HOF is the same as DHOG. HOF makes the sum of modulus value of light flow's horizontal component and vertical component as the amplitude of the light flow and makes the arctangent value of the ratio of vertical and horizontal component as the direction angle of the light flow. Then the light flow direction will be projected into the adjacent bin which is based on the amplitude of light flow. Considering that the light flow amplitude is small (less than a certain threshold), we will add a zero bin in the HOF feature which is used to hold the number of pixels. HOF descriptor also is normalized with the L_2 -norm. The other parameters are the same as the DHOG. So the dimensionality of HOF is $2 \times 2 \times 9 \times 3 = 108$.

2.3.3. Motion Boundary Histograms. Motion boundary histograms (MBH) are proposed by Yao et al. for detection of the human in 2014 [23]. It computes derivatives separately for the horizontal and vertical components. MBH is the gradient of light flow; it can get rid of uniform motion

TABLE 1: The number of each type of video.

Video	A2C	A3C	A4C	A5C	PLA	PSAB	PSAM	PSAP	Sum
Number	45	31	36	7	42	11	39	17	228
Patients	9	6	7	1	8	2	8	4	45
Normal	36	25	29	6	34	9	31	13	183

and keep the change of optical flow field. In a certain extent, it is a complement to the DHOG and HOF that is removing the camera movement to the human. The DHOG and HOF are generated from the optical flow, which makes the computational complexity of MBH small.

2.4. The Selection of Feature Encoding and Classifier. In this paper, we use the Fisher vector [24, 25] with the improved 256 visual words of GMM (signed square-rooting followed by $L2$ -norm) for features encoding. Fisher’s vector seen as the expansion of the BOV word bag model is the intermediate expression of image. Because Fisher’s vector fuses the advantages of production and is discriminant of Fisher’s kernel architecture, it can not only reflect the frequency of each visual word but also encode the differences of local feature in visual words information. In addition, it can represent the richer image feature compared to the word bag model. The high dimensional feature that combined with simple and effective linear classifier can achieve good classification effect. So, in this paper, we use the linear SVM classifier, which has the excellent performance of machine learning in classification and generalization. So linear SVM classifier has a lot of advantages compared to other learning algorithms in practice and is simple and convenient to use. Its performance is better in classification.

3. Experimental Results and Analysis

3.1. Data. In this experiment, 228 different echocardiography videos are collected from 72 different patients (including 58 normal individuals and 14 patients with cardiac wall motion abnormalities); these data were provided by Tsinghua University First Hospital. All the length of echocardiography video is 1 to 2 seconds and the videos are stored as 434×636 resolution and in 26-frame DICOM format (Digital Imaging and Communications in Medicine). Table 1 shows the detailed number of eight kinds of video.

3.2. Experiment Setup and Result Analysis. In this experiment the dense sampling grid of feature points is selected with 5 pixels in the process of dense sampling, and it can guarantee the best result. Then the tracking frame length is set to 15 frames in the process of tracking of feature points. In the process of motion boundary description, the default parameters are $N = 32$, $n_\sigma = 2$, and $n_r = 3$ in our experiment, and these parameters have the best effect through the experiments testing. In terms of selecting the feature encoding and classifier, this experiment adopts vlfeat-0.9.19 software for processing (download from <http://www.vlfeat.org/>). For the selection of training video and testing video, we use the

method of “half and half” (the total echocardiography video data are randomly selected, half as the training video and the other half as the testing video). And we analyze its classified results.

We use the confusion matrix to show the classification accuracy rate of the eight kinds of echocardiography video in Table 2, and the calculation method of accuracy rate is showed in function 3.1. Overall, the average classification accuracy of all the classes is 77.12%. We can find that the classification accuracy for each class is different. In particular, the classification accuracy for these classes, such as A5C, PSAB, and PSAP, is lower than the others. The main reason is shortness of experiment data corresponding to these three kinds of echocardiography videos. It leads to lack of training samples:

$$\text{accuracy rate} = \frac{\text{number of correct videos}}{\text{number of videos}}. \quad (6)$$

In view of the above experimental result that the classification accuracy can be affected by the number of the sample videos, in the following experiment, we cut off the less experimental data of three kinds of video (A5C, PSAB, and PSAP) and make further experiment with the rest of the five kinds of experimental data and compare the results (Table 3).

From Table 3, comparing with the eight kinds of echocardiography video and five kinds of the echocardiography video, we can find that the classification accuracy shows a significant improvement. The classification accuracy is promoted from original 77.12% to 98.36%.

Our experiments also make a classification test for echocardiography video of the primal three kinds of locations (AA, PLA, and PSA), its average classification accuracy is 100%, and it can satisfy the basic needs of medical echocardiography video classification.

Finally, compared with the other echocardiography video classification algorithm which is previously proposed our experimental result is increased a lot (illustrated in Table 4). And for the echocardiography video of the three kinds of primal locations, our experiment has the best accuracy. Therefore, the method of echocardiography video’s classification based on dense trajectory tracking and DHOG has an obvious improvement in accuracy compared with the previous algorithm.

4. Summary and Discussion

This paper is based on the classification of the echocardiography video with the dense trajectory tracking and DHOG algorithm. Firstly, we use the dense trajectory tracking and DHOG algorithm to describe the feature in each kind of echocardiography video and get the features and then apply the VLfeat software for Fisher vector to encode the features. Finally we use the linear SVM to get the classification results. From the experimental results, while the method adopted in this paper is compared with the previous method which is based on the 3D-SIFT echocardiography video classification, the classification accuracy is improved significantly. But the final average classification accuracy is still not high because

TABLE 2: Confusion matrix for eight kinds of echocardiography video classification.

	Classified result								Accuracy	
	A2C	A3C	A4C	A5C	PLA	PSAB	PSAM	PSAP		
Class	A2C	22	1	0	0	0	0	0	0	0.96
	A3C	1	15	0	0	0	0	0	0	0.95
	A4C	0	0	17	1	0	0	0	0	0.95
	A5C	2	0	1	0	0	0	0	0	0.00
	PLA	0	0	0	0	21	0	0	0	1.00
	PSAB	0	0	0	0	1	3	0	2	0.50
	PSAM	0	0	0	0	0	0	20	0	1.00
	PSAP	0	0	0	0	2	0	0	7	0.81

TABLE 3: Confusion matrix for five kinds of echocardiography video classification.

	Classified result					Accuracy	
	A2C	A3C	A4C	PLA	PSAM		
Class	A2C	22	1	0	0	0	0.96
	A3C	1	15	0	0	0	0.95
	A4C	0	0	18	0	0	1.00
	PLA	0	0	0	21	0	1.00
	PSAM	0	0	0	0	20	1.00

TABLE 4: The algorithm accuracy compared with other algorithms in this paper.

	The comparison of result	
	Classification accuracy of eight classes	Classification accuracy of three classes
[8]	72%	90%
[12]	70.4%	86%
[13]	74%	91%
[14]	76%	93%
[15]	76.5%	93%
[16]	75%	92%
[17]	73.2%	95%
[18]	71.3%	94%
Our method	77.12%	100%

of the lack of the echocardiography video data. So, in this paper, we get rid of the small amount of three kinds of experimental data and make the classification experiment again. After solving the problem of the sample data, the average classification accuracy of the rest of five classes of data compared to the average classification accuracy of eight classes has increased nearly 20.9%. And, in three classes of primal position of echocardiography video, the average classification accuracy is 100%. Therefore, the method of this paper can preliminarily achieve the requirement of the medical echocardiography video classification accuracy. For dealing with the classification of echocardiography video, the challenge is not just about the video noise caused by the low resolution of the video but also how to balance the computational complexity and time cost. In the future work,

we will gradually increase the database of eight kinds of echocardiography video and adopt or propose newer and more effective method to improve the classification accuracy and reduce the required time of the classification as well.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Fund in China under Grants nos. 61471124 and 61473090. Also this research forms part of WIDTH project that is financially funded by EC under FP7 programme with Grant no. PIRSES-GA-2010-269124.

References

- [1] R. Kumar, F. Wang, D. Beymer, and T. Syeda-Mahmood, "Cardiac disease detection from echocardiogram using edge filtered scale-invariant motion features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '10)*, pp. 162–169, IEEE, San Francisco, Calif, USA, June 2010.
- [2] S. K. Zhou, J. H. Park, B. Georgescu, C. Simopoulos, J. Otsuki, and D. Comaniciu, "Image-based multiclass boosting and echocardiographic view classification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, vol. 2, pp. 1559–1565, New York, NY, USA, June 2006.
- [3] D. Beymer, T. Syeda-Mahmood, and F. Wang, "Exploiting spatio-temporal information for view recognition in cardiac echo videos," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '08)*, pp. 1–8, IEEE, Anchorage, Alaska, USA, June 2008.
- [4] A. Shalhaf, H. Behnam, Z. Alizade-Sani, and M. Shojafard, "Automatic classification of left ventricular regional wall motion abnormalities in echocardiography images using nonrigid image registration," *Journal of Digital Imaging*, vol. 26, no. 5, pp. 909–919, 2013.
- [5] F. Sera, T. S. Kato, M. Farr et al., "Left ventricular longitudinal strain by speckle-tracking echocardiography is associated with treatment-requiring cardiac allograft rejection," *Journal of Cardiac Failure*, vol. 20, no. 5, pp. 359–364, 2014.

- [6] J. Y. Wang, Y. Li, Y. Zhang et al., “Bag-of-features based medical image retrieval via multiple assignment and visual words weighting,” *IEEE Transactions on Medical Imaging*, vol. 30, no. 11, pp. 1996–2011, 2011.
- [7] Y. Guo, Y. Wang, D. Kong, and X. Shu, “Automatic classification of intracardiac tumor and thrombi in echocardiography based on sparse representation,” *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 2, pp. 601–611, 2015.
- [8] Y. Qian, L. Wang, C. Wang, and X. Gao, “The synergy of 3D SIFT and sparse codes for classification of viewpoints from echocardiogram video,” in *Medical Content-Based Retrieval for Clinical Decision Support: Third MICCAI International Workshop, MCBR-CDS 2012, Nice, France, October 1, 2012, Revised Selected Papers*, vol. 7723 of *Lecture Notes In Computer Science*, pp. 68–79, Springer, Berlin, Germany, 2013.
- [9] M. Rap and A. Chacko, “Optimising the use of transoesophageal echocardiography in diagnosing suspected infective endocarditis,” *Acta Cardiologica*, vol. 70, no. 4, pp. 487–491, 2015.
- [10] N. Espinola-Zavaleta, M. E. Soto, A. Romero-Gonzalez et al., “Prevalence of congenital heart disease and pulmonary hypertension in Down’s syndrome: an echocardiographic study,” *Journal of Cardiovascular Ultrasound*, vol. 23, no. 2, pp. 72–77, 2015.
- [11] H. Wang, A. Kläser, C. Schmid, and C.-L. Liu, “Dense trajectories and motion boundary descriptors for action recognition,” *International Journal of Computer Vision*, vol. 103, no. 1, pp. 60–79, 2013.
- [12] I. Laptev, M. Marszałek, C. Schmid, and B. Rozenfeld, “Learning realistic human actions from movies,” in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’08)*, pp. 1–8, Anchorage, Alaska, USA, June 2008.
- [13] J. Liu, J. Luo, and M. Shah, “Recognizing realistic actions from videos in the wild,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR ’09)*, pp. 1996–2003, IEEE, Miami, Fla, USA, June 2009.
- [14] A. Gilbert, J. Illingworth, and R. Bowden, “Fast realistic multi-action recognition using mined dense spatio-temporal features,” in *Proceedings of the 12th International Conference on Computer Vision*, pp. 925–931, IEEE, Kyoto, Japan, October 2009.
- [15] W. Brendel and S. Todorovic, “Activities as time series of human postures,” in *Computer Vision—ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5–11, 2010, Proceedings, Part II*, vol. 6312, pp. 721–734, Springer, Berlin, Germany, 2010.
- [16] A. Kovashka and K. Grauman, “Learning a hierarchy of discriminative space-time neighborhood features for human action recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR ’10)*, pp. 2046–2053, San Francisco, Calif, USA, June 2010.
- [17] Q. V. Le, W. Y. Zou, S. Y. Yeung, and A. Y. Ng, “Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’11)*, pp. 3361–3368, June 2011.
- [18] B. Li, M. Ayazoglu, T. Mao, O. I. Camps, and M. Sznaiar, “Activity recognition using dynamic subspace angles,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’11)*, pp. 3193–3200, IEEE, Providence, RI, USA, June 2011.
- [19] G. C. Zhang and P. A. Vela, “Good features to track for visual SLAM,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’15)*, pp. 1373–1382, IEEE, Boston, Mass, USA, June 2015.
- [20] H. Wang and C. Schmid, “Action recognition with improved trajectories,” in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV ’13)*, pp. 3551–3558, Sydney, Australia, December 2013.
- [21] X. Li, G. Cui, W. Yi, and L. Kong, “A fast maneuvering target motion parameters estimation algorithm based on ACCF,” *IEEE Signal Processing Letters*, vol. 22, no. 3, pp. 270–274, 2015.
- [22] Y. Ji, M. Yang, Z. Lu, and C. Wang, “Integrating visual selective attention model with HOG features for traffic light detection and recognition,” in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV ’15)*, pp. 280–285, Seoul, Republic of Korea, June–July 2015.
- [23] B. Z. Yao, B. X. Nie, Z. Liu, and S.-C. Zhu, “Animated pose templates for modeling and detecting human actions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 436–452, 2014.
- [24] D. Veitch and P. Tune, “Optimal skampling for the flow size distribution,” *IEEE Transactions on Information Theory*, vol. 61, no. 6, pp. 3075–3099, 2015.
- [25] B. Klein, G. Lev, G. Sadeh, and L. Wolf, “Associating neural word embeddings with deep image representations using Fisher Vectors,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’15)*, pp. 4437–4446, IEEE, Boston, Mass, USA, June 2015.



Hindawi
Submit your manuscripts at
<http://www.hindawi.com>

