International Journal of Academic Research in Computer Engineering
Print ISSN: 2476-7638 and Online ISSN: 2538-2411
Vol. 2, No. 1, Pages. 40-45, May 2018

International Journal of Academic Research in Computer Engineering
Volume 2, Number 1, Pages 40-45, May 2018
www.ijarce.org
IJARCE

Detection of Normal and Abnormal Human Activities for Video Surveillance

Amir Karbalaei Hassan[1*]

1. Department of Computer Engineering, Urmia Branch, Islamic Azad University, Urmia, Iran.

Receive Date 2017.09.12; Accepted Date: 2018.02.17, Published Date: 2018.05.15

*Corresponding Author: A.Karbalaei Hassan (Amir.karbalaii@gmail.com)

Abstract
As the number of users of surveillance cameras increased in public places, the need was felt that the camera automatically recognize normal and abnormal activities were to quickly achieve higher detection and diagnosis without fatigue and human error can notify the protective forces. For this purpose, we introduce an integrated descriptor where the background and foreground information simultaneously apply Fourier transform to extract the information, we do feature extraction using Gabor filter to reduce the dimensions of a space and feature selection Maximum Relevance Minimum Redundancy (MRMR) use and the training and support vector machine classifier will recognize the activity. The advantages of this approach are the property of not separating foreground from background to meet the needs of Arbitration in Sport games as well and the use of classifiers using feature selection step and the speed and accuracy MRMR to reduce the dimensions noted.

Keywords: Activity Detection, Video Descriptor, Dimension Reduction, Feature Extraction, Monitoring

1. Introduction

Control by surveillance sites such as stadiums, banks… In addition to being a great human power is needed due to fatigue or human errors could be of error. Therefore, a system that can intelligently and wrong are far from ordinary activities and extraordinary security forces to identify and notify in recent years the scientific community has been highly regarded. In studies all human activity detection systems to follow these 4 steps (1-download video 2-preprocessing 3-detecting human 4-activity detection) several studies in the literature and for the detection of human activity in the third and fourth stages competed.

Then extract information for easier movement, background and foreground simultaneously to the frequency domain we find that this challenge is a response to the need for monitoring systems for sports that require simultaneous background and foreground. Instead of the third stage of previous studies that human exposure was reduced in size by selecting features and convert do the space MRMR. In the fourth step, the activity detection using Support Vector Machine (*SVM)* step by step, accuracy, and speed up the activity we detected. The purpose of this research is finding algorithms that accurately and more quickly than humans can distinguish normal and abnormal behaviors or have any individuals or collective activity detections. In such a system after data pre-processing that is performed is ready to apply the learning stage model step learning model, the order of the data pre-processing of data that is selected according to the mining method identified and generated model To assess transferred to the next stage of evaluation and interpretation model. Since this system for classifying normal or abnormal activities or behavior detection is used in different environments, different environments and data-bank definition of normal and abnormal behavior will be different for training. In the second part of this paper research on the subject presented in the third part of the proposed article will be reviewed. In the fourth section, we evaluate the proposed method,

and in fifth section conclusions and recommendations for future work will form.

2. Related Works

As regulatory systems, human-machine interaction is fertile ground for the rise, in recent decades Activities in the field of interaction between humans and machines has increased dramatically. Detection systems proposed activities followed the four steps in preprocessing, on reducing the optimal range evaluated using background subtraction, and luminous flux is doing. A large number of modeling background methods by Krystany and et al. in 2010 and Alhabyan in 2008 has been developed [2]. Find distance measured between successive frames of an image sequence, are called luminous flux. The first method for calculating the variable optical flux in the image sequence was introduced in 1981 by Horn et al. [4]. The most crucial and important part of the system, detecting human body position is estimated. The section is divided into two main sections commonly-used technique, component-based techniques and analysis to identify single window [3].

Finally, both diagnostic methods and activities can be classified in two categories of single-layer and hierarchical methods. Monolayer methods, more suitable for diagnostic and hierarchical methods used to detect activities [7]. In 2010, Forrest et al. [9] using one-class support vector machines article entitled discovery event-driven monitoring of their video streams. In 2014, Chen and et al. [6] a new algorithm based on the network to identify human activity on video by learning SVM offered, as well as in 2014, Kang and et al. based on this classification [5] the discovery of unusual behavior by operating a hybrid presented in crowded scenes. In 2016 [11] Yu et al. Markov model-based, non-supervisory activity detection algorithm was introduced. In 2015, Omer and et al. [1] using the graphics and video analysis, an article entitled Analysis of Space-Time Video to explore their disorder. All procedures provided in the past, requires estimation methods were based on the background and foreground were detected and acted on the basis of luminous flux caused by human movement. In cases, the total pre-processing stage, revealing the background, and finally do a proper classification algorithm, is very consuming in time and memory. As well as in some activities such as sports arbitration in the information stage and move hand or foot simultaneously need that with using previous methods, such as background removal and separate extraction of the body situation did not miss this important information.

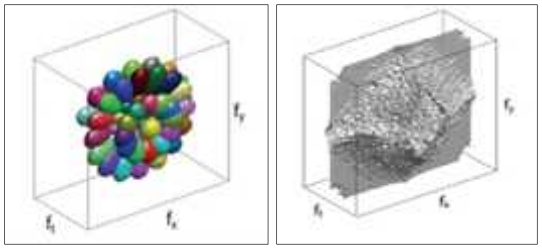
3. Proposed Method

The proposed methods in the past, requires estimation of background and foreground were detected. In this case, the total pre-processing stage, revealing the background and result in appropriate classification algorithm, is very consuming in time and memory. Therefore, a model of situations that require separate extraction of the body in different frames and background and foreground are not separate from each other can improve system performance and reduce the lead time required. The separation of the foreground not the background while extracting feature is also useful in sports arbitration acts. This requires some adverse conditions mentioned above prompted us to design a video integrated descriptor that scene data and move simultaneously extracted and then applied to positively reduce the feature size given that our accuracy is not reduced, at this point we are faced with increasing accuracy of the proposed method shows that this point of differentiation.

At first we cut three-dimensional video frames of the three sections in our tests. For computing less and that the elections do not affect the performance of more than three slices on the further improvement. Cut video in the frequency domain to the frequency domain we identify the easier it is. At this point it is assumed that the camera is in a fixed location. The background information is identical in all video frames. However, because convolution in the primary field is the same as multiplying the frequency domain and because we have the background information can move easily separate background information in the frequency domain. For transmission of video frames used in the frequency domain of the discrete Fourier transform. Three-dimensional Fourier transform *f (x, y, t)* on the space and time is calculated as follows:

$$F(f_x, f_y, f_t) = \frac{1}{MNT}\sum_{x=0}^{M}\sum_{y=0}^{N}\sum_{t=0}^{T} f(x,y,t)e^{-2®\left(\frac{xf_x}{M}+\frac{yf_y}{N}+\frac{tf_t}{T}\right)} \quad (1)$$

M, *N* and *T*, respectively, width, height and duration of the video is cut and x and y and t spatial position and time anywhere in the volume is created. Three-dimensional Gabor filter used for extracting features from the bank. Gabor filter is modeled structure of the human eye and a direction is appropriate descriptor. This filter edges in different directions and movement of an object or describe it well and display filters see in Figure 1.



(A) (B)

Figure 1. (A) 68 three-dimensional Gabor filter (B) Applying Gabor filters bank 50 UCF

Three-dimensional transfer function of each filter, in accordance with a spatial frequency f_{r0} along marked with polar angles $\bar{}_0$ and $¤_0$ direction and in a spherical coordinate system can be expressed as follows:

$$G(f_r,\bar{},¤) = exp\{-\frac{(f_r - f_{r0})^2}{2\pm_r^2} - \frac{(\bar{} - \bar{}_0)^2}{2\pm_{\bar{}}^2} - \frac{(¤ - ¤_0)^2}{2\pm_¤^2} \quad (2)$$

That:

$$¤ = arccos\left(\frac{f_z}{\sqrt{f_x^2 + f_y^2 + f_t^2}}\right)$$

$$\bar{} = arctan\left(\frac{f_y}{f_x}\right)$$

$$f_r = \sqrt{f_x^2 + f_y^2 + f_t^2}$$

Parameters $\pm_r$, $\pm\bar{}$ and $\pm_¤$, respectively, radial and angular bandwidth that stretch spatio-temporal filter in the frequency domain signals. Filtering effect of frequency spectrum in Figure 2 you'll see.

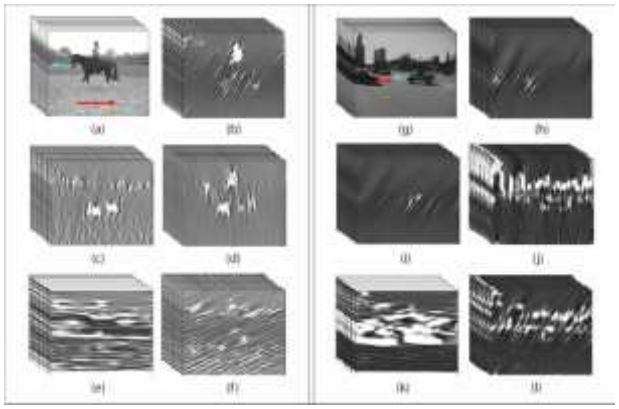


Figure 2. Filtering effect of frequency spectrum

In the fourth stage of the three-dimensional inverse discrete Fourier transform is applied, features are extracted in the frequency domain and should return to the primary domain.

$$\sum_{u=0}^{M-1}\sum_{V=0}^{N-1}\sum_{t=0}^{T-1} F(F_x,f_y,f_t) e^{-2®\left(\frac{xf_x}{M}+\frac{yf_y}{N}+\frac{tf_t}{T}\right)} F(x,y,t) = \frac{1}{MNT} \quad (3)$$

In the fifth round of the volumes produced per cutting and average those to the next step (inset) send. For the following information on each cut volumes with the help they need to teach our system.

In the sixth step below average volumes obtained from the previous step are concatenated. The next stage is required during production, reduced feature vectors. At this stage, because the feature vector is generated 104,448 so we need to cut later. To reduce the dimension of feature vectors obtained, First, we applied Principal Component Analysis (PCA) dimension reduction with this method but did not satisfy us, so we used the feature selection *MRMR* then we applied to *MRMR* applied *PCA* to see if it comes to dimensions less with higher accuracy? The results of the experiments will answer the questions posed.

MRMR feature selection method based on the relationship between features and concept (or a label) c is defined. In the "most relevant" try to approximate the relationship between selected features and most of the glue that's exactly what "most dependent" also was looking for. The difference is in how to calculate *D*, so that all information on the method *D* on average common features of *Xi* and Class *C* is defined labels. This relationship is defined by the following mathematical form:

$$MaxD(S,c), \qquad D = \frac{1}{|S|}\sum_{X_i \in S} I(x_i;c) \tag{4}$$

However, although the most relevant characteristics are obtained in this way, but there may be two features that rely heavily on their time and power categorized by the elimination of one of them does not change or that change is very minor and negligible. Here the second phase of the previous phase is added in order to minimize redundancy features, which in addition to their close relationship with the concept to (c), against each other with mutual exclusion, i.e. the relationship do not have a strong bond with each other, in mathematical terms, we can say:

$$MinR(S), \qquad R = \frac{1}{|S|^2}\sum_{X_i, X_j \in S} I(x_i;x_j) \tag{5}$$

Two upper limits in the vicinity of the measures called up communication and defines the minimum redundancy is called for short *MRMR*, *φ(D,R)* operator used to combine two phases have been proposed with the aim of optimizing the *R, D* simultaneously and the form is defined as follows:

$$Max\Phi = (D,R),\ \Phi = D - R \tag{6}$$

One of the obvious advantages *MRMR* method than the "most dependent" probability density estimate is that in *MRMR* of multiplicity $p(x_1,x_2,\ldots x_m)$ and $p(x_1,x_2,\ldots x_m,c)$ is to avoid and Instead, the joint probability density $p(x_i, x_j)$ and $p(x_i,c)$ taking the task easier and more is possible. Finally, the *SVM* hierarchical classification and diagnosis will be treated. *SVM* applied to any data in the first round scored at least one label for the second stage of the probability that the first stage of the tagged data were used to new entries to the *SVM* as the second stage at this point the data has not label the first stage of our label.

4. Test Results

We have found this method to test the performance of two data sets *50UCF* [13] and *UCSD* [12] which we used publicly available. 50UCF data collection including videos on the web, in uncontrolled conditions from 50 different activity categories with more than 100 video is for each category as well as data collection *UCSD*, which includes pedestrian with the vehicle and is without a vehicle.

For all tests, we used three video cutting. Select more than three video cut does not affect further improve performance, length feature vector in our tests, which is 104448, 68 filters, 512 sub-volumes and 3 have been cut resulting video (104448 = 512 × 68 × 3). In Table 1 as a result of feature selection for reduced dimensions and compare the accuracy of our results.

Table 1. Compare 4 methods due to reduced dimensions and maximum amount of precision used in any 2 banks

Number the highest accuracy in UCF50	Most of the accuracy of the proposed method UCF50 by repeating 3	Reduced dimensions in the bank UCF50	Number highest precision at UCSD	The greatest amount of accuracy by repeating UCSD 30	Reduced dimensions in the bank UCSD	Method
3	0/6873	6681* 104448	1	1	70 *104448	All Data
1	0/5667	6681* 13336	8	0/5750	70 *70	All Data +PCA
2	0/6740	6681* 6357	29	1	70* 18695	MRMR
1	0/5350	6681* 1336	8	0/5750	70 *70	MRMR +PCA

By comparing the accuracy of the plot (3) and (4) with four proposed methods is applied after first reducing the *PCA* method we apply the reduced dimensions. But after this reduction did not satisfied us, so we used the screening method of feature selection. One of these four methods which consists of a feature selection also applies to its *PCA*, in rows of Table 1. These four methods with regard to the accuracy of the frequency of occurrence of 30 in the bank *UCSD*, 70 samples each were compared. As the number of samples less than the number of samples *50UCF* the bank, but the number of repetitions Bank *50UCF*, that it was acceptable samples were taken 3. In Table 1, we see them.

A) According to the second column of Table least amount of trash after the bank *UCSD* applied to the *PCA* is concerned that both methods will be 70*70. *PCA* applied to the data that makes it features dimensions of not more samples. *PCA* apply to reduce the data size. But much more important data lost in this way, on the other hand to reduce *MRMR* acceptable after we arrived and the amount of repeat accuracy of 1 in 29.

B) According to the fifth column of the table that the greatest reduction in the bank *50UCF* the *PCA* will be applied to unblock and, as noted above in this way we would lose a lot of information after they *PCA* methods to be applied, *MRMR* least extent with the highest precision 6740/0 will be compared to other methods that occurred in the second iteration.

C) As in Table 1 were *MRMR* method in *UCSD* Bank and the Bank *50UCF* highest accuracy *ALL DATA* will be after the procedure. This result by screening random feature selection in the properties went very excellent. The two charts below are four methods applied to reduce bank size for both *UCSD* and *50UCF* according to the number of repeats for both banks in the form of (3) and Figure (4) viewing.

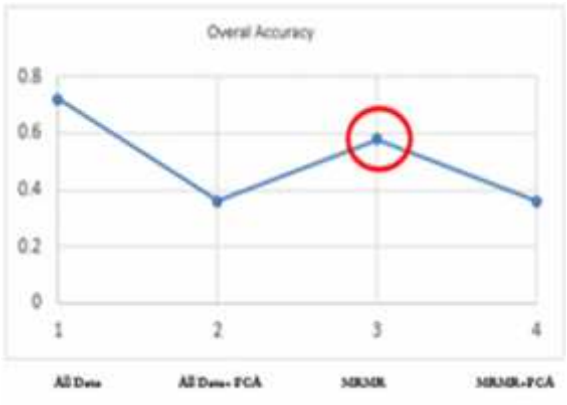


Figure 3. Average accuracy of SVM in 4 methods of the 30 runs

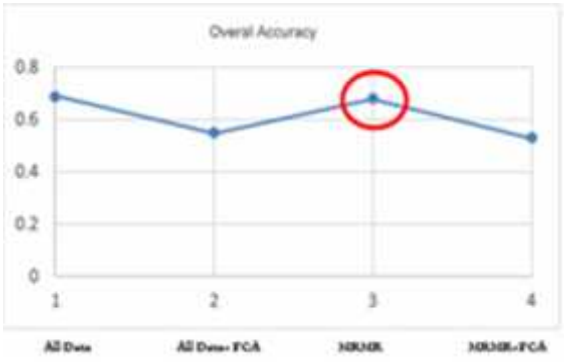


Figure 4. Average accuracy of SVM on 50 UCF graph in different ways

In order to evaluate the performance of an integrated three-dimensional descriptor, making it popular descriptors Generalized Integrated Search Tree (GIST) [8] and Spatio-Temporal Interest Point (STIP) [10] on 50UCF Compare dataset UCSD and we, as Figure (5) is shown. We have carefully descriptor 40/67% in the bank *50UCF* on 50 categories of activity in this series that looks much better than other descriptors. As well as in Figure (6) has shown. The proposed method for *UCSD* Bank, which has 70 samples of normal and abnormal film, has an accuracy of 100% over the same banks that its accuracy is higher than the two descriptors.

Fig. 5: Compare the diagnostic accuracy than in GIST and STIP activities proposed in the bank 50UCF

Figure 6. Compare the diagnostic accuracy than in GIST and STIP activities proposed in the bank UCSD

5. Conclusions and Future Work

The system design is always around fatigue, confusion, short, optimal processing power at the same time and can be quite intelligent human activity in sensitive environments put under surveillance and out of the ordinary detect unusual activity and to notify security forces can bring better security to people, therefore, in this article we integrated a video descriptor for use in detecting unusual activity in which the man introduced after the cut videos we applied Fourier them to move data into the frequency domain,

where they proceed to extract the information, Gabor filter is applied to spectrum produced and performed feature extraction. Then, after reduction with feature selection and SVM training MRMR and turning space and we recognize the behavior. According to the results MRMR the first position in terms of the precision of the cut is later. Among the pre-processing of the data, you can choose to select a subset of features and characteristics noted. that in the midst of feature selection methods, MRMR that constituted the foundation of this article, can be combined with other feature selection methods like Joint Mutual Information (JMI) and Mutual Information-based Feature Selection (MIFS) and MRMR & MIFS and PCA applied again to each of them, the actions each of which can increase the accuracy and speed of diagnosis and recommendations for future work that will be more cuts later.

We use SVM in the learning stage, but we can learn if a more accurate classification of the data may also be desirable. Other learning algorithms that can be used to increase the accuracy of the classification proposed: decision tree based methods, methods of rule-based, memory-based reasoning, neural networks, and methods based on the theory, support vector machines. in addition to better categorize that we did not label any data without combining classifiers are used that one of the purposes of this article constituted for this purpose, hierarchical *SVM* used at all stages. Here can be applied after the first stage, the third in the later stages of other classification algorithms such as neural networks or decision trees and rule-based methods… Also be used. The innovative aspect of the proposed system can reduce the dimension of feature selection *SVM* and *MRMR* hierarchical separation of the foreground not the background while noting feature extraction.

Reference

[1] B. Antic, B. Ommer, Spatio-temporal Video Parsing for Abnormality Detection, Computer Vision and Pattern Recognition, pp. 1-15, 2015.

[2] Sh. Y. Elhabian, Kh. M. El-Sayed, S. H. Ahmed, Moving object detection in spatial domian using background removal techniques-state-of-art, Recent Patents on computer, Science, Vol. 1, pp. 32-54, 2008.

[3] D. Gavrial, The visual analysis of human movement: A Survey. Computer Vision and Image Understanding, Vol. 73, pp. 82-98, 1999.

[4] B. Horn, B. Schunck, Determining optical flow, Artificial Intelligence, 17, pp. 185-204, 1981.

[5] S. HyunCho, H. BongKang, Abnormal behavior detection using hybrid agents incrowded scenes, Pattern Recognition Letters, 44, pp. 64-70, 2014.

[6] W. Lin, Y. Chen, J. Wu, H. Wang, B. Sheng, H. Li, A new Network-Based Algorithm for Human Activity Recognition in Videos, IEEE Trans actions on circuits and Systems for Video Technology, 24(5), 2014.

[7] J. Liu, J. Luo, M. Shah, Recognizing realistic actions from videos in the wild, IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[8] O. Oliva, A. Torralba, G. Dugue, J. Herault, Global semantic classification of scenes using power spectrum templates, Challenge of Image Retrieval, pp. 1-12, 1999.

[9] C. Piciarelli, G. Foresti, Surveillance- oriented event detection in video streams, IEEE Intell. Syst, 26(3), pp. 32-41, May/Jun 2010.

[10] H. Wang, M. Ullah, A. Klaser, I. Laptev, C. Schmid, Evaluation of local spatio-termporal features for action recognition, In British Machine Vision Conference, 2009

[11] D. Yeo, B. Han, J. Han, Unsupervised Co-Activity Detection from Multiple Videos Using Absorbing Markov Chain, Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), 2016.

[12] www.svcl.ucsd.edu/projects/anomaly/UCSD_Anomaly_Dataset.tar.gz.

[13] www.crcv.ucf.deu/data/UCF50php.