International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.9 (2015), pp.339-352
http://dx.doi.org/10.14257/ijmue.2015.10.9.35

Abnormal Event Detection Based on Saliency Information

Zhijun Fang[1, 2], Fengchang Fei[1], Yuming Fang[1*], Lei Shu[1] and Wanggen Wan[2]

[1]*School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, Jiangxi, China*

[2]*School of Communication and Information Engineering, Shanghai University, Shanghai, China*

[1]*zjfang@gmail.com*

[*]*Corresponding author: Yuming Fang, E-mail: fa0001ng@e.ntu.edu.sg*

Abstract

Abnormal event detection is a challenging task in video analysis. In this paper, we propose a new abnormal event detection algorithm for surveillance videos. It is well accepted that human eyes are extremely sensitive to abnormal events and they can quickly pay attention to the locations of these abnormal events in visual scenes. Thus, the characteristics of the Human Visual System (HVS) can be used for abnormal event detection. By exploiting the characteristics of the HVS, we propose an abnormal event detection algorithm based on saliency information. Firstly, the saliency information is extracted from video frames based on the feature contrast. The motion information of video frames is calculated by the multi-scale histogram optical flow (MHOF). Based on the features of saliency information and MHOF, the Support Vector Machine (SVM) is used to train and predict the abnormal events in visual scenes. Experimental results show that the proposed abnormal event detection method can obtain much better performance than the existing ones over the public video database.

Keywords: *abnormal event detection; crowd analysis; saliency information; Support Vector Machine*

1. Introduction

Abnormal event detection in crowded scenes is an important task in intelligent surveillance video systems. It refers to detect and respond to the abnormal changes or behaviors of humans or objects in videos. There are various abnormal detection algorithms in crowded scenes such as abnormal crowd behavior detection [1-3], human abnormal action detection [4], traffic incident detection [5], *etc.*

One type of abnormal event detection methods is designed based on object detection and tracking [6]. These methods first focus on the moving objects in visual scenes, and then track the object trajectory. Generally, the trajectories of moving objects are different from each other. However, there are general rules to track the motion of different objects. When abnormal events occur, almost all the trajectories of moving objects would be influenced, by which we can identify abnormal events. During this process, the challenging task is how to locate and track the objects. And there are various factors leading to failure or performance of abnormal event detection. For example, in crowded scenes, human tracking is difficult to implement and the computational complexity is high.

Existing studies also try to detect abnormal events in video frames by extracting features from temporal and spatial domains [7]. In that type of methods [7], the global and local

ISSN: 1975-0080 IJMUE

features are extracted to be combined for the final abnormal event detection. The local feature is extracted by the differences between the target and its surrounding regions. The global feature is extracted from visual scenes globally without taking into account specific targets in visual scenes. Generally, the features used in this type of method include mixtures of dynamic textures [8], global cues [9], social Force Model [10], Histogram of Optical Flow (HOF) [11], saliency feature [12], Multi-scale histogram of optical flow [13], global optical flow orientation histogram [14], and so on.

2. Related Work

2.1. Saliency Information Extraction

Visual attention is an important mechanism of the Human Visual System (HVS) and the earliest study on human eye's attention started in 1890. When observers look at a scene, they would pay attention to the salient regions in the scene. This illustrates that in the HVS, selective attention would mainly process visual information of salient regions in visual scenes.

Human eyes can quickly and effectively focus on important events through the mechanism of selective attention in complex scenes. Human vision can always pay attention to the salient regions, which are different from their neighboring regions. Generally, the abnormal event existing in video frames can be represented by the sudden change in spatial or temporal dimension in video frames. Thus, the abnormal event can also be considered as salient regions in video frames. In this paper, we present an abnormal event detection algorithm based on saliency information. Here, we use the saliency detection model in [15] to extract the salient regions in video frames. The video frame is firstly divided into image patches, and then the saliency value of each patch is determined according to the difference between one patch and all other image patches from the features of color, intensity, and orientation as follows [15]:

$$s_i = \sum_{i \neq j} w_{i,j} d_{i,j} \tag{1}$$

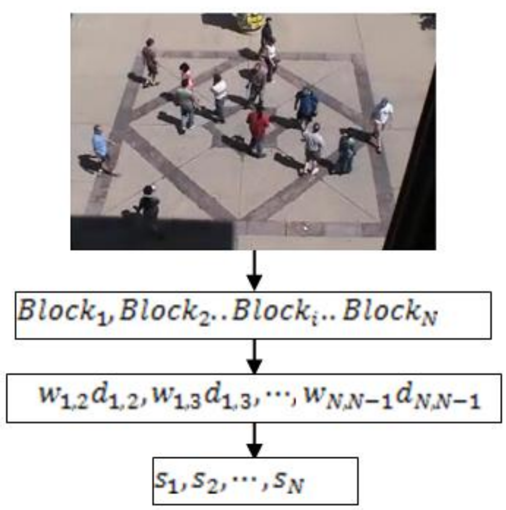
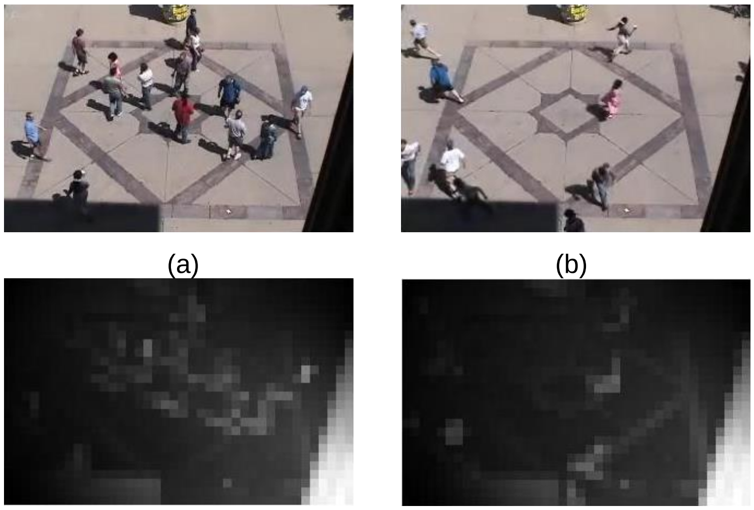


Figure 1. Framework of the Saliency Information Extraction

Where s_i is the saliency value of image patch i, $d_{i,j}$ is the difference between patches i and j, $w_{i,j}$ is the weight between patches i and j, determined by the human visual sensitivity. In [15], $d_{i,j}$ is represented by the difference between amplitude spectrum of patches i and j, the framework is shown in Figure 1. Figure 2 shows the saliency information of a normal frame and an abnormal frame.



(a) (b)

(c) (d)

Figure 2. (a) Normal Frame; (b) Abnormal Frame; (c), (d) Saliency Maps of Images (a) and (b)

2.2. Multi-Scale Histogram of Optical Flow

Optical flow is represented by the 2D instantaneous velocity of all pixels in an image. The 2D velocity vector is the point of the 3D velocity vector in the projection imaging surface. Thus, optical flow not only includes the motion information of the observed object, but also contains the information about the 3D structure of the scene. Each pixel (i, j) has a 2D velocity vector $\left(o_{i,j}^{x}, o_{i,j}^{y}\right)$. If the abnormal event detection task takes optical flow of each pixel as a feature, the computational complexity will be high, and the pixel noise will also influence the results. To obtain the better performance, multi-scale histogram of optical flow (MHOF) is proposed in [13]. In MHOF, it first divides each video frame into small image patches, and then each patch is classified into 16 classes. By using the histogram of 16 classes as the patch features instead of optical flow, that method greatly reduces the amount of computation, and achieves the effect of suppressing the noise in the optical flow.

MHOF preserves more precise motion information than the traditional histogram of optical flow (HOF). In the study [13], MHOF can better describe the current frame scene changes, and can detect abnormal events in video sequence accurately.

The framework of the proposed MHOF is show in Figure 3. First of all, every video frame is divided into image patches with the same size $M \times M$. Then, we calculate the optical flow matrix $O(\boldsymbol{o}^x, \boldsymbol{o}^y)$ of each patch, and the MHOF of each patch. Equations (2) and (3) are used to calculate each pixel class-label $class_{i,j}$, where $\left(o_{i,j}^{x}, o_{i,j}^{y}\right)$ is the optical flow of pixel (i, j), θ is the angle of $o_{i,j}^{x}, o_{i,j}^{y}$.

$$c_{i,j} \in \begin{cases} 0 & \left\| o_{i,j}^{x}, o_{i,j}^{y} \right\| \le tr \\ 1 & \left\| o_{i,j}^{x}, o_{i,j}^{y} \right\| > tr \end{cases} \quad (2)$$

(tr is the magnitude threshold)

$$class_{i,j} = round\left(\theta\left(o_{i,j}^{x}, o_{i,j}^{y}\right)/(\pi/4)\right) + 8 \times c_{i,j} \quad (3)$$

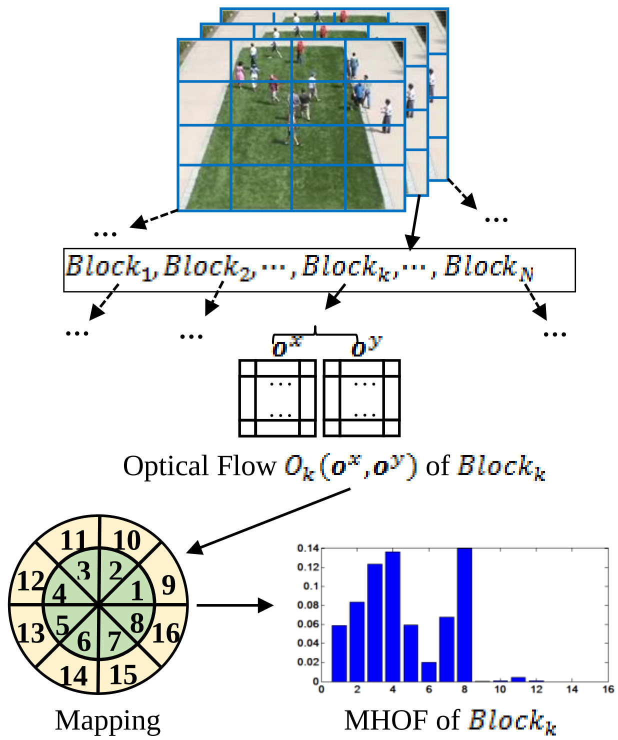


Figure 3. Framework of MHOF

2.3. SVM

Support vector machines (SVM) [16] is widely used in statistical classification and regression analysis in various applications [17]. It for a two-class separable learning task, the samples are mapped into a high dimensional space, and in this space a hyper-plane will classify the samples into two classes. To obtain promising classification performance on the data, we should select the optimal hyper-plane. Thus, two side-planes are set up and they are parallel to the hyper-plane and have the same distances to the hyper-plane. The distance between the two side-planes is called margin. The hyper-plane is in the middle of these two side-planes. The larger the margin, the higher the accuracy of the classifier is. Thus the purpose of the SVM algorism is to obtain the maximum marginal hyper-plane.

There is an example of two-class linearly separable learning task in Figure 4. There are many sample points in a 2-dimension Descartes coordinates and there are two coordinate values for each sample point x and y. In 2-D space, the representation of the hyper-plane is a straight line. These sample points have two classes: positive and negative. There is a hyper-plane in Figure 4 (a) and (b). The aim of SVM is to find a maximum marginal hyper-plane (MMH), and thus the hyper-plane of Figure 4(a) is the final result of SVM. To calculate the MMH, we assume that the classification function is as follows:

$$f(x) = \boldsymbol{\omega}^T \boldsymbol{x} + b \tag{4}$$

Where $\boldsymbol{\omega}$ is the normal vector of MMH, $\{(\boldsymbol{x}, \boldsymbol{y})\}$ is the sample set. The hyper-plane can be defined as

$$\boldsymbol{\omega}^T \boldsymbol{x} + b = 0 \tag{5}$$

The two side-plane equation is

$$\begin{cases} \boldsymbol{\omega}^T \boldsymbol{x} + b = -1 \\ \boldsymbol{\omega}^T \boldsymbol{x} + b = +1 \end{cases} \tag{6}$$

Figure 4(a), (b) illustrates the three plane.

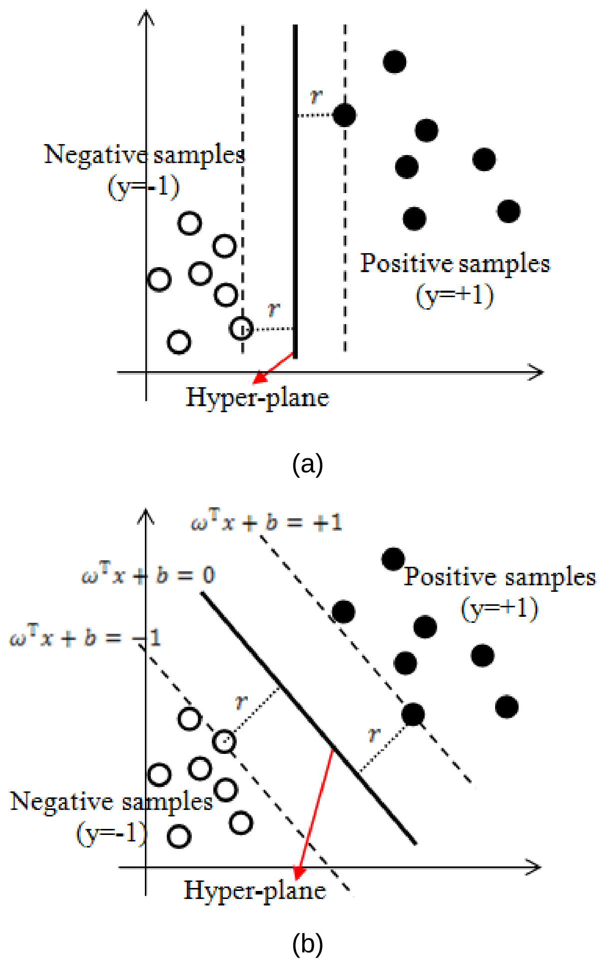


Figure 4. The Samples of the Hyper-Plane for a Linearly Separable Case

According to the point to plane distance formula, we can get:

$$r = \frac{\omega^T x + b}{\|\omega\|} = \frac{f(x)}{\|\omega\|} \tag{7}$$

In order to eliminate the orientation, $\boldsymbol{y}$ should be multiplied with r.

$$\tilde{r} = yr = \frac{y(\omega^T x+b)}{\|\omega\|} = \frac{yf(x)}{\|\omega\|} \tag{8}$$

If we fix the functional margin to be equal to 1, then the given sample points $\{\boldsymbol{x}_i, y_i\}$ satisfy the follows

$$\begin{cases} \boldsymbol{\omega}^T \boldsymbol{x}_i + b \geq 1 & for \quad y_i = +1 \\ \boldsymbol{\omega}^T \boldsymbol{x}_i + b \leq -1 & for \quad y_i = -1 \end{cases} \tag{9}$$

The particular points $\{\boldsymbol{x}_i, y_i\}$ which can make the equalities of Equation (9) to be satisfied are called supper vectors. So from Equation (8), the distance from supper vectors $\boldsymbol{x}^*$ to the hyper-plane is

$$r^* = \frac{y^* f(x^*)}{\|\omega\|} = \frac{y^*}{\|\omega\|} \tag{10}$$

If r^* is max, the hyper-plane is the optimal hyper-plane (MMH). So r^* should be maximized with respect to $\boldsymbol{\omega}$ and b.

$$\begin{gathered} max \frac{1}{\|\omega\|} \\ s.t.\ y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1, i = 1,2 \ldots, n \end{gathered} \tag{11}$$

The final solution of $\boldsymbol{\omega}$ and b construct the MMH classifier, also known as SVM classifier.

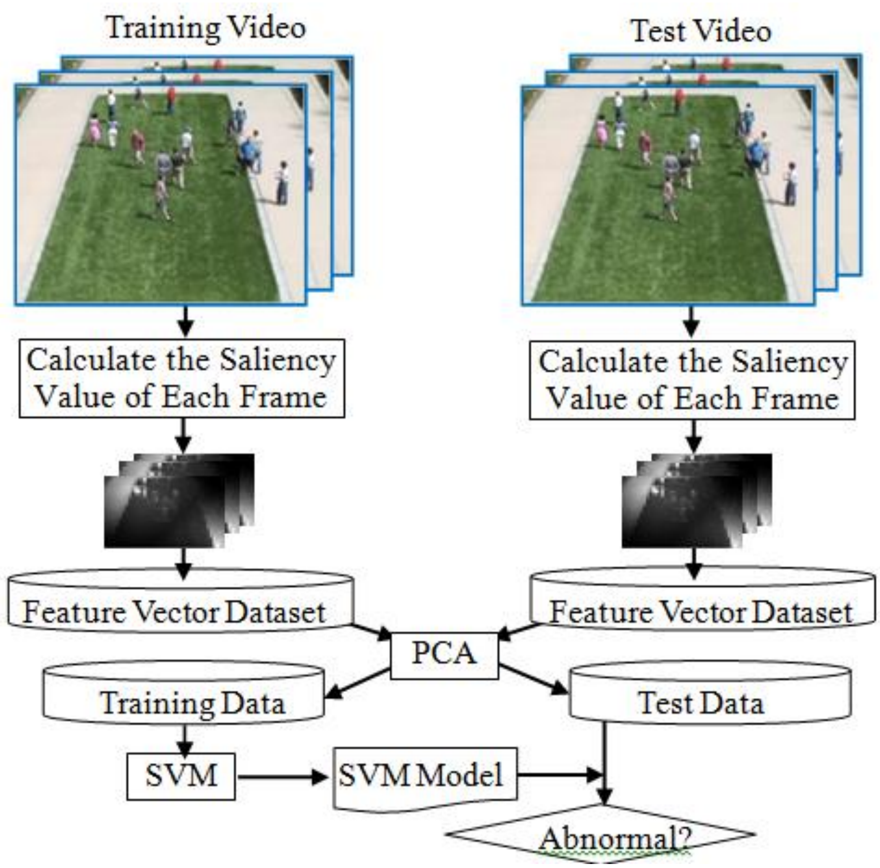


Figure 5. Framework of the Proposed SI Model

3. The proposed Method

3.1. Abnormal Event Detection

Saliency information can reflect the characteristics of human vision, thus in this paper we build a model of abnormal event detection based on saliency information (SI model), which is illustrated in Figure 5. The algorithm is implemented as follows:

Step1 Divide each training video frame into m image patches.

Step2 The saliency value of training video $\boldsymbol{S}_{train}\{\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_i, \cdots, \boldsymbol{s}_n\}$ can be calculated, where n is the total number of video frames in training video; $\boldsymbol{s}_i = (s_1, s_2, \cdots, s_j, \cdots s_m)$ is the saliency value matrix of frame i, and s_j is the saliency value of patch j.

Step3 Use PCA to transform $\boldsymbol{S}_{train}$ into $PCA\_\boldsymbol{S}_{train}$ to reduce the dimensions of $\boldsymbol{S}_{train}$.

Step4 Train the PCA_S_{train} with the SVM and obtain the corresponding prediction model.

Step5 According to step1~step2, we can calculate the saliency value vector of each test frame $\boldsymbol{s}_{test}(s_1, s_2, \cdots, s_j, \cdots, s_m)$, and reduce its dimensions to $PCA\_\boldsymbol{S}_{test}$.

Step6 Detect $PCA\_\boldsymbol{S}_{test}$ using SVMmodel, to check whether the test frame is abnormal.

3.2. Spatio-Temporal Abnormal Event Detection Model

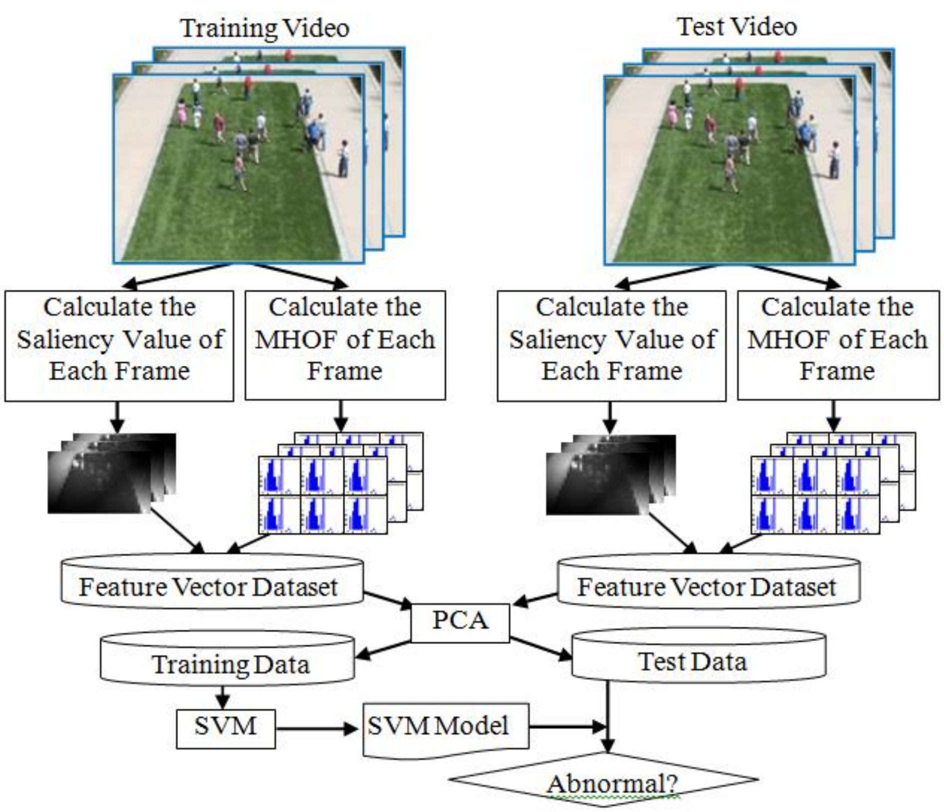


Figure 6. Framework of the Proposed SI+MHOF Model

As indicated previously, saliency information (SI) represent the important information in visual scenes. Optical flow is the velocity vector of pixels, and thus we use MHOF to extract temporal features of video. Based on the above analysis, we combine SI and MHOF to build a spatio-temporal model of abnormal detection (SI+MHOF model), as illustrated in Figure 6.

Step1 Divide each training video frame into m image patches.

Step2 The saliency value matrix of training video $\boldsymbol{S}_{train}\{\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_i,\cdots,\boldsymbol{s}_n\}$ can be calculated; where n is the total frame number of training video, $\boldsymbol{s}_i=\{s_1,s_2,\cdots,s_j,\cdots s_m\}$ is the saliency value matrix of frame i, and s_j is the saliency value of patch j.

Step3 According to Figure 3 and Equations (2) (3), we calculate the MHOF matrix of training video $\boldsymbol{H}_{train}\{H_1,H_2,\cdots,H_i,\cdots,H_n\}$. $H_i=\{\boldsymbol{h}_1,\boldsymbol{h}_2,\cdots,\boldsymbol{h}_j,\cdots \boldsymbol{h}_t\}$ is the MHOF matrix of frame j and $\boldsymbol{h}_j$ is the MHOF vector of patch j.

Step4 by taking $(\boldsymbol{s}_i,H_i)$ as the features of frame i, we can get the training data set $data_{train}\{(\boldsymbol{s}_1,H_1),(\boldsymbol{s}_2,H_2),\cdots,(\boldsymbol{s}_n,H_n)\}$.

Step5 Use PCA to transform $data_{train}$ into PCA_data_{train} to reduce the dimension of $data_{train}$.

Step6 Train the PCA_data_{train} with a SVM to obtain the corresponding SVM model.

Step7 According to step1~step4, the $data_{test}(\boldsymbol{s}_k,H_k)$ of each test frame can be calculated, and it can be transformed into PCA_data_{test}.

Step8 Detect PCA_S_{test} using SVM model, to determine whether the test frame is abnormal.

4. Experiment

There are three different scenes in UMN dataset [18]: lawn scene, indoor scene, plaza scene and 7738 frames in total. There are 11 abnormal events in 7739 frames.

In the experiment, we compare SI, SI+MHOF and MHOF algorithm. In [13] for calculating MHOF, each frame is divided into 20 image patches and has 320 features in total. In order to ensure consistent feature dimension, in SI model, each frame is divided into 320 image patches and thus obtains 320 saliency values, SI+MHOF model is the algorithm by combining these two features, and thus, each frame has 640 features.

In this paper, we use FPR-TPR curve as the evaluation method. As shown in Figure 7, the x-axis is called the false positive rate (FPR), called precision. The y-axis is the true positive rate (TPR), called recall. FPR and TPR are calculated as follow:

$$FPR=\frac{False\ positive}{False\ positive+True\ negative}$$
$$TPR=\frac{True\ positive}{True\ positive+False\ negative} \tag{12}$$

The ROC area under the curve (AUC) can be used to provide the overall performance evaluation. With larger AUC, the better detection results are.

We randomly select training frames from experimental frames with a certain proportion, the remaining part is used as the test. In order to ensure accurate results, the experimental results are average of 20 test results. In the following subsections, we will separately analyze the experimental results of these three scenes.

4.1. Lawn Scene

There are 1453 frames in this video sequence. The frames with large-change pedestrian motion are labeled as abnormal frames. So there are 1333 frames to be labeled as normal frames, and 120 frames are labeled as abnormal frames. Figure 7 illustrates the curve of FPR-TPR with the different proportion of training samples. AUC values are shown in Table 1.

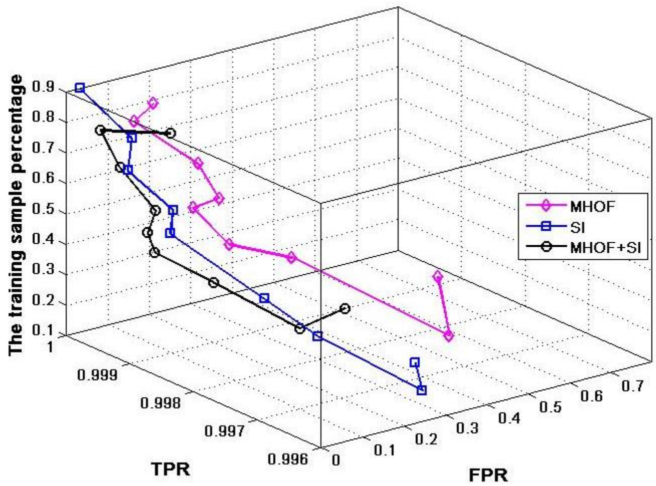


Figure 7. The FPR-TPR Curves of Lawn Scene

Table 1. The AUCs of Lawn Scene

The training samples percentage	MHOF AUC	SI AUC	SI+MHOF AUC
0.1	0.6157	**0.7800**	0.7428
0.2	0.7594	**0.8747**	0.8514
0.3	0.7975	**0.9136**	0.8988
0.4	0.8753	0.9424	**0.9531**
0.5	0.8927	0.9550	**0.9716**
0.6	0.9085	0.9644	**0.9776**
0.7	0.9173	**0.98332**	0.98197
0.8	0.9464	0.9776	**0.9876**
0.9	0.9533	0.9834	**0.9876**

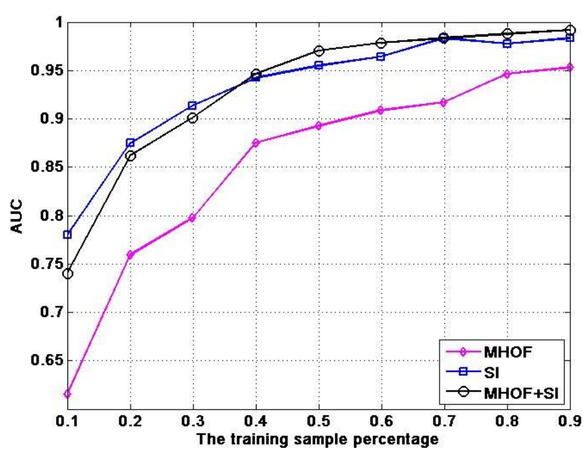


Figure 8. The AUC Curves of Lawn Scene

With the difference of proportion of training samples, the AUC of these three algorithms have the different values. The corresponding curves are shown in Figure 8.

From Table 1 and Figure 8, with the different proportion of training samples, the results of SI and MHOF+SI are better than MHOF algorithm, and the results of SI and MHOF+SI are similar. However, with the increasing of training samples, the results of MHOF+SI are better than SI.

4.2. Indoor Scene

There are 4144 frames in this video sequence. The frames with large-change pedestrian motion are labeled as abnormal frames. So there are 3715 frames to be labeled as normal frames, and 429 frames are labeled as abnormal frames. Figure 9 illustrates the curve of FPR-TPR with different proportions of training samples. AUC values are shown in Table 2. With the difference of proportion of training samples, the AUC of the three algorithms have the different values. The corresponding curves are shown in Figure 10.

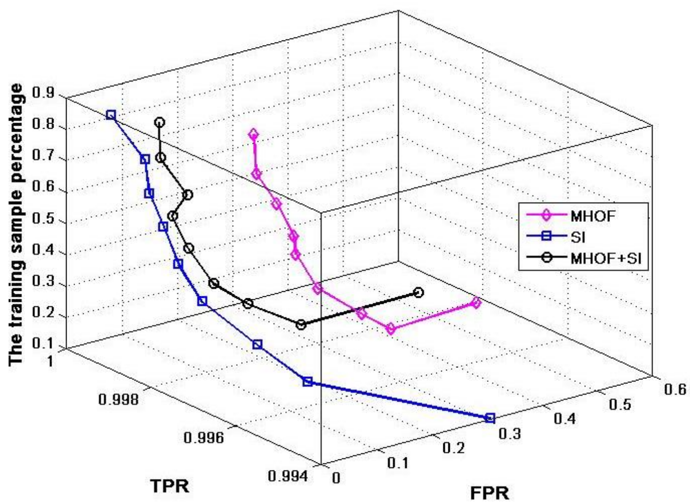


Figure 9. The FPR-TPR Curves of Indoor Scene

Table 2. The AUCs of Indoor Scene

The training samples percentage	MHOF AUC	SI AUC	SI+MHOF AUC
0.1	0.7064	**0.8426**	0.7315
0.2	0.8281	**0.9409**	0.8700
0.3	0.8644	**0.9627**	0.9203
0.4	0.8971	**0.9800**	0.9479
0.5	0.9056	**0.9864**	0.9592
0.6	0.9244	**0.9874**	0.9648
0.7	0.9321	**0.9928**	0.9728
0.8	0.9457	**0.9927**	0.9813
0.9	0.9377	**0.9956**	0.9804

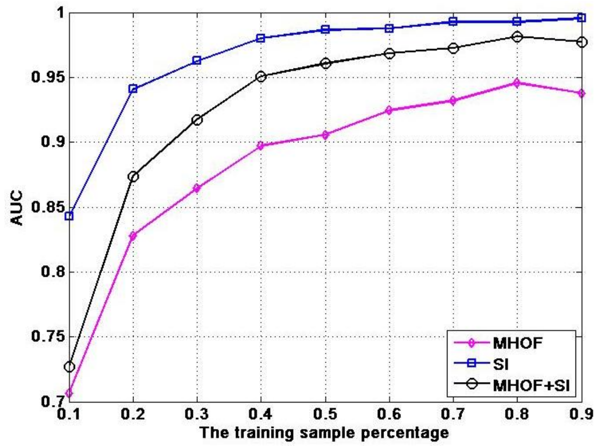


Figure 10. The AUC Curves of Indoor Scene

From Table 2 and Figure 10, with different proportion of training samples, the results of SI and MHOF+SI are better than MHOF algorithm, and the results of SI are better than MHOF+SI. This may be due to the act that MHOF features affect the accuracy.

4.3. Plaza Scene

There are 2142 frames for this video sequence. The frames with large-pedestrian motion are labeled as abnormal frames. So there are 1974 frames to be labeled as normal frames, and 168 frames are labeled as abnormal frames. Figure 11 illustrates the curve of FPR-TPR with different proportions of training samples. The AUC values are shown in Table 3. With difference proportions of training samples, the AUC of these three algorithms have different values, and the corresponding curves are shown in Figure 12.

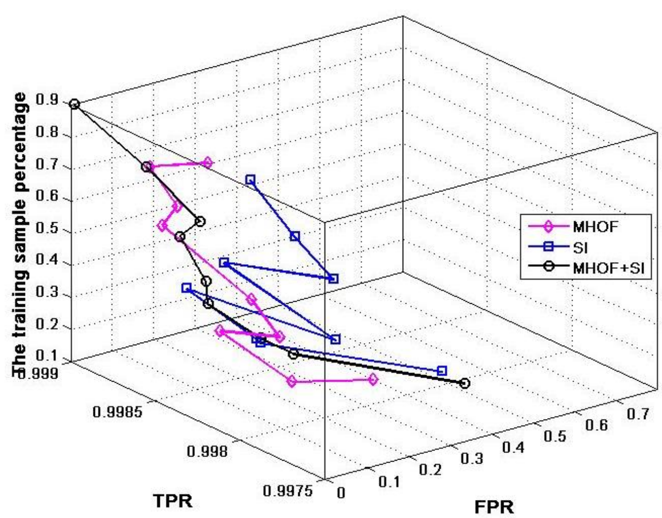


Figure 11. The FPR-TPR Curves of Plaza Scene

Table 3. The AUCs of Plaza Scene

The training samples percentage	MHOF AUC	SI AUC	SI+MHOF AUC
0.1	**0.7992**	0.7327	0.7303
0.2	**0.9159**	0.8769	0.8732
0.3	**0.9411**	0.9262	0.9266
0.4	0.9575	0.9486	**0.9597**
0.5	0.9703	**0.9730**	0.9716
0.6	0.9685	**0.9832**	0.9728
0.7	0.9783	0.9806	**0.9852**
0.8	0.9872	0.9921	**0.9932**
0.9	0.9890	0.9857	**0.9947**

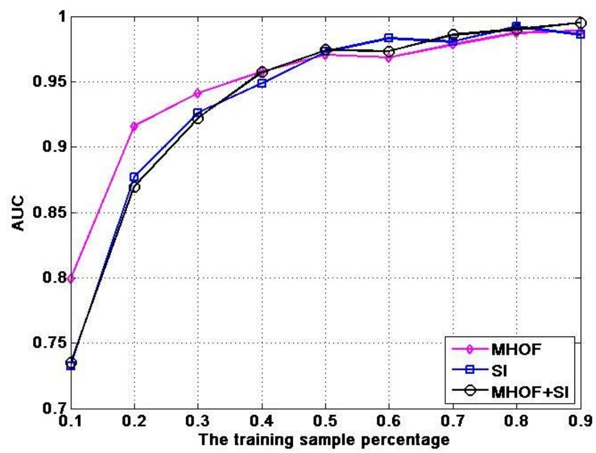


Figure 12. The AUC curves of Plaza Scene

From Table 3 and Figure12, when the percentage of training samples is less than 0.4, the result of MHOF is better than SI and MHOF+SI. If the percentage is bigger than 0.4, the results of MHOF+SI and SI are better than MHOF.

5. Conclusion

We propose a method of abnormal event detection based on saliency information. We combine SI and MHOF features together to design a novel spatio-temporal abnormal event detection model, which consider spatial and temporal characteristics of the video. Experimental results show that in most cases the results of SI and MHOF+SI are better than MHOF.

Acknowledgements

This work was partially supported by Postdoctoral Science Foundation of China (No. 2013M541508), the Funds from DoE of Jiangxi Province (No.GJJ14347, GJJ14318), and NSF of Jiangxi Province (No. 20142BAB217011, 20141BDH80003).

References

[1] S. H. Cho and H. B. Kang, “Integrated multiple behavior models for abnormal crowd behavior Detection”, IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI); Santa Fe, NM, April 22-24, **(2012)**.
[2] L. Wang and M. Dong, “Real-time detection of abnormal crowd behavior using a matrix approximation-based approach”, IEEE International Conference on Image Processing (ICIP); Orlando, FL, September 30-October 3, **(2012)**.
[3] W. Y. Ren, G. H. Li, J. Chen and H. Z. Liang. Abnormal crowd behavior detection using behavior entropy model. International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), **(2012)** July 15-17; Xian, China
[4] P. Cui, F. Wang, L. F. Sun, J. W. Zhang and S. Q. Yang, “A matrix-based approach to unsupervised human action categorization”, IEEE Transactions on Multimedia, vol.14, **(2012)**.
[5] J. Yuan, “Discriminative video pattern search for efficient action detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, **(2011)**.
[6] T. Hassner, Y. Itcher and K. G. Orit, “Violent flows: real-time detection of violent crowd behavior”, 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); Providence, RI, June **(2012)**.
[7] Benezeth Y., Jodoin P., Saligrama V. and Rosenberger C., “Abnormal events detection based on spatio-temporal co-occurences”, Computer Vision and Pattern Recognition (CVPR); Miami, FL, USA, June20-25, **(2009)**.
[8] V. Mahadevan, W. Li, V. Bhalodia and N. Vasconcelos, “Anomaly detection in crowded scenes”, Computer Vision and Pattern Recognition (CVPR); San Francisco, CA, June 13-18, **(2010)**.
[9] R. Ma, L. Li, W. Huang and Q. Tian, “On pixel count based crowd density estimation for visual surveillance”, 2004 IEEE Conference on Cybernetics and Intelligent Systems, vol. 1, December 1-3, **(2004)**.
[10] R. Mehran, A. Oyama and M. S, “Abnormal crowd behavior detection using social force model”, Computer Vision and Pattern Recognition (CVPR); Miami, FL, June20-25, **(2009)**.
[11] A. Adam, E. Rivlin, I. Shimshoni and D. Reinitz,”Robust real-time unusual event detection using multiple fixed-location monitors”, IEEE Transactions on Pvattern Analysis and Machine Intelligence, vol. 30, **(2008)**.
[12] V. Gopalakrishnan, Y. Hu and D. Rajan, “Salient region detection by modeling distributions of color and orientation”, IEEE Transactions on Multimedia, Vol. 11, **(2009)**.
[13] Y. Cong, J. S. Yuan and J. Liu, “Abnormal event detection in crowded scenes using sparse representation”, Pattern Recognition, vol. 46, **(2013)**.
[14] T. Wang and H. C. Snoussi, “Detection of abnormal visual events via global optical flow orientation histogram”, IEEE Transactions on Information Forensics and Security, vol. 9, **(2014)**.
[15] Y. Fang, W. Lin and B. S. Lee, “Bottom-up saliency detection model based on human visual sensitivity and amplitude spectrum”, IEEE Transactions on Multimedia, vol. 14, **(2012)**.
[16] V. N. Vapnik, “The Nature of Statistical Learning Theory”, Springer-Verlag, New York, USA, **(1995)**.
[17] V. N. Vapnik, “Statistical Learning Theory”, Wiley-Interscience Publication, USA, **(1998)**.
[18] http://mha.cs.umn.edu/proj_events.shtml

Authors

Zhijun Fang, received the Ph.D. degree from Shanghai Jiaotong University, Shanghai, China. He is currently a professor in the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. His research interests include image processing, video coding, and pattern recognition. He was the General Chair of HHME2013 (the 9th Joint Conference on Harmonious Human Machine Environment) and the General Co-Chairs of ISITC2014 (2014 International Symposium on Information Technology Convergence).

Fengchang Fei is a PhD candidate in management science and engineering, Jiangxi University of Finance and Economics, Nanchang, China. His research interests include video processing and machine vision.

Yuming Fang is currently a faculty member in the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. He received the Ph.D. degree in Computer Engineering from Nanyang Technological University, Singapore in 2013. He was a (visiting) Postdoc Research Fellow in IRCCyN lab, PolyTech' Nantes & Univ. Nantes, Nantes, France, University of Waterloo, Waterloo, Canada and Nanyang Technological University, Singapore. His research interests include visual attention modeling, visual quality assessment, image retargeting, computer vision, 3D multimedia processing, *etc.*

Lei Shu is a PhD candidate in management science and engineering, Jiangxi University of Finance and Economics, Nanchang, China. His research interests include image processing and machine vision.

Wanggen Wan received his PhD degree from Xidian University, China in 1992. From 1991 to 1992, he was a visiting scholar in Computer Engineering Dept. of Minsk Radio Engineering Institute, former USSR. Since 2004, he has been with the School of Communication and Information Engineering, Shanghai University where he is currently a Full Professor, Dean of Academic Affairs, Director of Institute of Smart City, and Program Leader of Circuit and System. Prof. Wan is a fellow of IET, and IEEE Senior Member. He has been Co-Chairman for many international conferences since 2008. His research interests include computer graphics, signal processing and data mining. He is a coauthor of approximately 200 academic papers.