Effective Scheme for Global Abnormal Event Detection for Surveillance Video

Fangxu Dong[1,2,3] and Dong Hu[1,2,3]
[1]Education Ministry's Key Lab of Broadband Wireless Communication and Sensor Network Technology
[2]Education Ministry's Engineering Research Center of Ubiquitous Network and Heath Service
[3]Jiangsu Province's Key Lab of Image Processing and Image Communications, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China

***Abstract*—An effective algorithm for global abnormal detection from surveillance video is proposed in this paper. The algorithm is based on sparse representation. To deal with the illumination change in video scenes, specific feature extract methods are designed for corresponding illumination conditions. In the case of non-uniform illumination, features are extracted directly on the original image; in the case of uniform illumination, features are extracted on the binary image obtained by threshold segmentation on the difference image, where the thresholds are computed by the Otsu's method. The features extracted on normal video are used to learn an over-complete dictionary. Then, the sparse reconstruction cost over the dictionary is used to detect abnormal events. Experiments on the open global abnormal dataset and the comparison to the state-of-the-art methods validate effectiveness and quickness of our algorithm.**

Keywords-abnormal detection; illumination change; binary images; sparse reconstruction cost

I. INTRODUCTION

The abnormal detection is one of the hot spots in visual surveillance in recent years and a variety of methods have been proposed. In [1]-[2], Histogram of optical flow (HOF) from image patches were computed on original image and foreground image, then, support vector machine or the Gaussian Process were adopted to detect abnormal events. In [3], Histogram of oriented gradient (HOG) on image patches were computed on the spatio-temporal video volumes, then the kernel based direct density ratio were estimated to judge the abnormal events. In [4], dynamic texture were used to detect abnormal events. In [5] spatio-temporal gradients on spatial temporal video volumes are computed, then the normal behavior models were learned by statistical learning methods to detect abnormal events. Sparse reconstruction cost (SRC) over the normal dictionary adopted to detect abnormal events has been proved to be an effective method in abnormal detection [6], since sparse representation is suitable to represent high-dimensional samples with less training data. However, they didn't consider the relationship between scene illumination change and sparse representation when detect abnormal events which resulted the scene feature not well-express. In this paper, we consider the influence of scene illumination change on the SRC abnormal detection method, employ different approaches to extract feature based on whether the scene illumination is stable. For non-uniform illumination scenes feature is extracted directly on the original image, while for uniform illumination scenes feature is extracted on the binary image which obtained by threshold segmentation on the difference image, which the threshold computed by the Otsu's method. Then we use the k-means singular value decomposition (KSVD) to learn dictionary on normal video feature. Finally, abnormal events are detected according the SRC [6] over the dictionary learned above. This greatly saves the abnormal detection time and reduces the complexity of abnormal detection algorithm.

The rest of the paper is organized as follows. In Section II, scene classification, difference image on gray image, binary image by threshold segment, the HOG feature on binary image, dictionary learning, sparse reconstruct on dictionary are introduced. According the SRC we can judge abnormal events. In Section III, we present the experiment results on real world video scenes to verify the validity of our algorithm. Finally, Section IV concludes the paper.

II. OUR METHOD

In this section, we present the details of the proposed algorithm. First, consider the illumination change, we classify the surveillance video into illumination uniform scenes and illumination non-uniform scenes by detecting the illumination uniformity of the video scenes. Then for illumination non-uniform scenes, we extract HOG feature directly on the original image; while in illumination uniform scenes, we use our proposed method to extract HOG feature: (1)transform the original images into gray images, if the images in video surveillance is gray images, this step could skip; (2)calculate the neighboring difference images; (3)adopt the Otsu's method to calculate the threshold on difference images; (4)obtain the binary images by employing threshold segmentation on difference images; (5)extract HOG feature on binary images. Next we use KSVD method to learn the normal over-complete dictionary from the HOG feature extracted above. Finally we reconstruct the input test feature by a sparse linear combination of an over-complete dictionary. Then the global abnormal events are detected by sparse reconstruction cost. The flowchart of our proposed algorithm is shown in Figure I.

A. Scene Classification

The video scene is classified by calculate the image brightness variance. The large image brightness variance in video scene which much possible is illumination non-uniform scene, while the small image brightness variance which much possible is illumination uniform scene.

B. Feature Extract

For illumination non-uniform scenes, we extract the HOG descriptors directly on the original images. For illumination uniform scenes, we extracting the HOG descriptors directly on the original images could not well express the scene appearance and dynamic. In this case we propose our effective feature extract method.

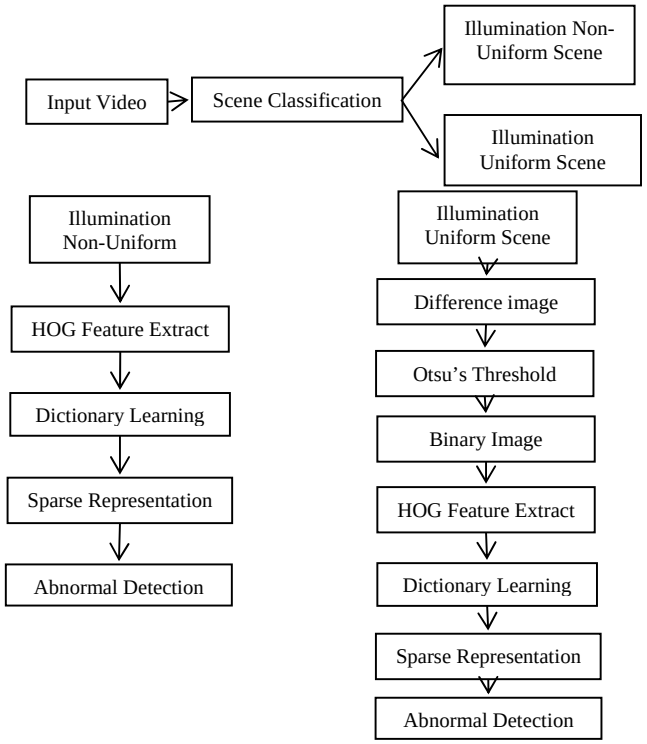


FIGURE I. THE FLOWCHART OF OUR PROPOSED ALGORITHM

Step 1: Transform the origin images f into gray images f_{gray}, f_{red}, f_{green} and f_{blue} denotes the red component, green component and blue component of the original image. The transform method is shown in Equ(1).

$$f_{gray} = 0.299 f_{red} + 0.587 f_{green} + 0.144 f_{blue} \quad (1)$$

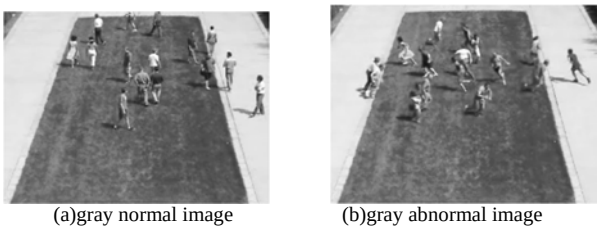
In this paper we adopt the images with single color for following feature extraction. If the images in video surveillance is already single color images, this step can be skipped. There in this paper we employ the gray component.

Step 2: Calculate the difference images $f_{graydiff}$ on the gray images which obtained via the difference between the gray pre-image f_t and the gray next-image f_{t+1}, f_t and f_{t+1} are shown in Figure II. The difference process is shown in Equ(2). The difference image f_{diff} are shown in Figure III.

$$f_{graydifff} = f_{t+1} - f_t \quad (2)$$

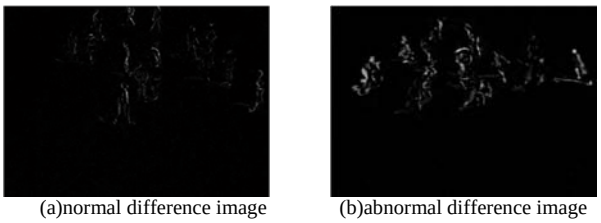
Step 3: Adopt the Otsu's method on the gray difference image to calculate the threshold T.

Step 4: Obtain the binary image by employing threshold T segment on difference image, if the gray value in difference image beyond T, the gray value may be set to 1, otherwise the gray value may be set to 0.



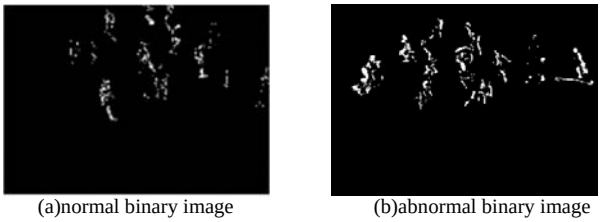
(a)gray normal image (b)gray abnormal image

FIGURE II. GRAY IMAGES



(a)normal difference image (b)abnormal difference image

FIGURE III. DIFFERENCE IMAGE



(a)normal binary image (b)abnormal binary image

FIGURE IV. BINARY IMAGE

Step 5: Extract HOG on the binary images obtained above. The binary image f_{bw} are shown in Figure IV.

C. Dictionary Learning

In this section, we address the problem of how to select the dictionary given an initial candidate HOG feature pool as $B = [b_1, b_2, \cdots, b_k] \in R_{m \times k}$, where m denotes the HOG descriptor dimension and each column vector $b_i \in R_m$ denotes a normal feature, $i = 1, 2, \cdots k$. Our goal is to find an over-complete dictionary $D_{m \times n}$ from B, where $n \ll k$. Such that the set B can be well reconstructed by D and the size of D is smaller than B. In this paper we adopt the KSVD algorithm to learn the dictionary. KSVD is an iterative method that alternates between sparse coding of the examples based on the current dictionary and a process of up dating the dictionary atoms to better fit the data. The over-complete dictionary matrix $D_{m \times n}$ that contains n signal-atoms for columns, $\{d_j\}_{j=1}^{n}$, a signal $y \in R_m$ can be represented as a sparse linear combination of these atoms. The representation of y may

either be exact $y = Dx$ or approximate $y \approx Dx$, satisfying $\|y - Dx\|_p \le \varepsilon$. Vector $x \in R_n$ contains the representation coefficients of the signal y. Our objective function is

$$\min_{D,X} \left\{ \|Y - DX\|_F^2 \right\} \quad \text{subject to} \quad \forall i, \|x_i\|_0 \le T_0 \tag{3}$$

First we consider the sparse coding step, where we assume that D is fixed, and consider the above optimization problem as a search for sparse representations with coefficients summarized in the matrix X. We use the Orthogonal Matching Pursuit (OMP) algorithms to solve this problem.

Second we turn to the process of updating the dictionary together with the nonzero coefficients. Assume that both D and X are fixed and we put in question only one column in the dictionary d_k and the coefficients that correspond to it, the k th row in X, denoted as x_T^k (this is not the vector x_k which is the k th column in X). Here, we employ SVD to find alternative d_k and x_T^k. With the d_k and x_T^k suggested above, we may now return to Equ(3) to calculate the optimal D.

D. Global Abnormal Event Detection

This section details how to determine a testing sample y to be normal or not. As we mentioned in the previous subsection, the features of a normal frame can be linearly constructed by only a few bases in the dictionary D while an abnormal frame cannot. A natural idea is to pursue a sparse representation and then use the reconstruction cost to judge the testing sample. In order to advance the accuracy of prediction, two more factors are considered here:

In practice, the deformation or any un-predicated situation may happen to the video. Motivated by [6], we extend the dictionary from D to $\Phi = [D, I_{m\times m}] \in R_{m\times(m+n)}$.

If a basis in the dictionary appears frequently in the training dataset, the cost to use it in the reconstruction must be lower, since it is a normal basis with high probability. Therefore, we design a weight matrix to capture this prior information. $W = diag[w_1, w_2, \cdots, w_n, 1, \cdots, 1] \in R_{(m+n)\times(m\times n)}$ Each $w_i \in [0,1]$ corresponds to the cost of the i th feature. For the artificial feature set $I_{m\times m}$ in our new dictionary Φ, the cost for each feature is set to 1.

Now, we are ready to formulate this sparse reforestation problem:

$$x^* = \arg\min_x \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda_1 \|Wx\|_1 \tag{4}$$

where $x = [x_o, e_o]^T$, $x_0 \in R_n, e_0 \in R_m$. This can be solved by linear programming using the interior-point method, which uses conjugate gradients algorithm to compute the optimized direction. Given a testing sample y, we design SRC using the minimal objective function value of Equ(5) to detect its abnormality:

$$S_w = \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda_1 \|Wx^*\|_1. \tag{5}$$

A high SRC value implies a high reconstruction cost and a high probability of being an abnormal sample.

E. Global Abnormal Event Detection

Normal or abnormal events usually occur during some number of successive frames. The abnormal events can be considered as false alarms if they just appear few frames intermittently in the long normal sequence, which can be adjusted to normal. Similarly, it also works on short clips of normal events which are found in long abnormal sequence. We know a high SRC value implies a high reconstruction cost and a high probability of being an abnormal sample. So we can post-process the detection results by presetting a threshold on the number of detected abnormal frames with high SRC. By analyzing the SRC of consecutive frames, we can twice detect the abnormal events. This post-processing makes our abnormal detection algorithm much more robust and much more accurate.

III. EXPERIMENTS

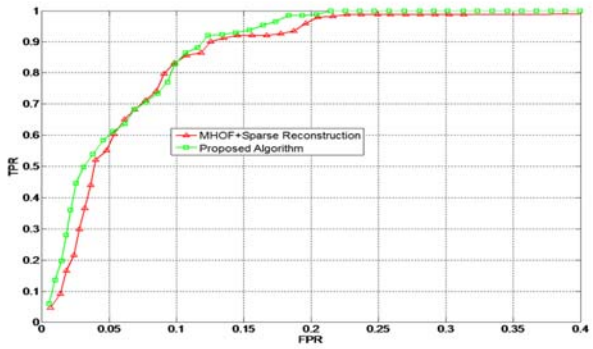
To validate the effectiveness and quickness of our proposed algorithm, we apply it on the open UMN dataset to test the global abnormal events. The UMN dataset consists of 3 different scenes of crowded escape events, and the total frame number is 7740(1450, 4415 and 2145 for scene 1, 2, 3, respectively) with a 320*240 resolution. Scene 1 and scene 3 are illumination uniform scenes, while scene 2 are illumination non-uniform scene. The normal events are pedestrians walking randomly on the square or in the mall, and the abnormal events are human spread running at the same time towards the same direction or all the direction. The receiver operating characteristic (ROC) curves by frame-level measurement of the detection results are used it to evaluate the accuracy of our algorithm. The results shows that our algorithm can obtain satisfactory detection performances as paper [6], but saves detection time than paper [6].

A. Illumination Non-Uniform Scene

For scene 2, HOG is extracted directly on the original image. The training dictionary is initialized from the first 300 frames in the scene and others is left for testing. We split each image into 3*3 sub-regions, for each sub-region we calculate a 9 bins HOG. So each image can be denoted as a 81 dimension vector. The normal/abnormal detection results are shown in Figure V. The detection time performance are shown in Table I.

(a)illumination non-uniform normal

(b)illumination non-uniform abnormal



(c)ROC curve of illumination non-uniform scene compared with paper [6]

FIGURE V. DETECTION RESULTS OF ILLUMINATION NON-UNIFORM SCENES

TABLE I. ILLUMINATION NON-UNIFORM SCENE

Method	Detection time(per frame)
paper[6]	0.0022s
Proposed method	0.0018s

Detection time is the abnormal detection time on every image in video streams. The results show that our algorithm satisfies the need of real time and costs less time than paper [6] because our algorithm uses lower feature dimension to detect abnormal events, but obtains satisfactory result.

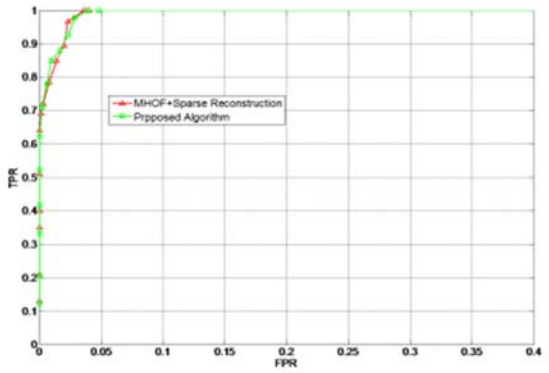
B. Illumination Uniform Scenes

For scenes 1 and 3, we extract HOG on binary images by adopting our proposed feature extract method. The normal/abnormal detection results are shown in Figure VI The detection time performance compared with paper [6] are shown in Table II and Table III.

For illumination uniform scenes, our method also achieves good results.

(a)illumination uniform lawn normal

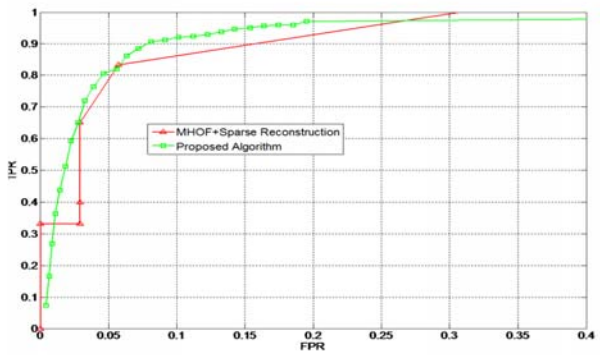
(b)illumination uniform lawn abnormal



(c)ROC curve of illumination uniform lawn scene compared with paper [6]

(d)illumination uniform plaza normal

(e)illumination uniform plaza abnormal



(f)ROC curve of illumination uniform plaza scene compared with paper [6]

FIGURE VI. DETECTION RESULTS OF ILLUMINATION UNIFORM SCENES

TABLE II. ILLUMINATION UNIFORM LAWN SCENE

Method	Detection time(per frame)
paper[6]	0.0018s
Proposed method	0.0015s

TABLE III. ILLUMINATION UNIFORM PLAZA SCENE

Method	Detection time(per frame)
paper[6]	0.0018s
Proposed method	0.0016s

IV. CONCLUSION

An effective scheme for global abnormal detection is proposed. The scheme is based on scene classification and sparse reconstruction over the normal dictionary. This scheme has been tested on several sequences and the results have shown that is effective to detect abnormal events. Future work

will aim at how to reduce the wrong detection rate and select our train-set more over-complete. Maybe we can train the dictionary online to suit the real environment to make our abnormal detection algorithm more robust.

ACKNOWLEDGMENT

This work was supported by the Project of the Priority Academic Program Development of Jiangsu Higher Education Institutions: Information and Communication Engineering.

REFERENCES

[1] Nannan Li, Xinyu Wu, Huiwen Guo, Dan Xu, Yongsheng Ou and Yenlun Chen. Anomaly Detection in Video Surveillance via Gaussian Process. International Journal of Pattern Recognition and Artificial Intelligence. Vol. 29, No. 6 (2015) 1555011 (25 pages).

[2] T. Wang, H. Snoussi, Detection of visual abnormal events via global optical flow orientation histogram. IEEE TRANSACTIONS ON INFORMATION AND SECURITY, Vol. 9, No. 6 (2014).

[3] YT Chen, WH Fang, CY Lee, KW Cheng, Abnormal Detection in Crowded Scenes via Kernel Based Direct Density Ratio Estimation, IEEE China Summit and International Conference on signal and information processing(2015), 15(1):99-104.

[4] Li Weixin, Vijay Mahadevan and Nuno Vasconcelos. Anomaly Detection and Localization in Crowded Scenes. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL.36, JANUARY 2014.

[5] Kai-Wen Cheng, Yie-Tarng Chen, and Wen-Hsien Fang, Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression, IEEE Conference on Computer Vision & Pattern Recognition (2015),pp:2909-2917.

[6] Yang Cong, Junsong Yuan, Ji Liu. Abnormal event detection in crowded scenes using sparse representation. Pattern Recognition. Volume.46, pp: 1851-1864, 2013.