Abnormal Event Detection based on Analysis of Movement Information of Video Sequence

Tian Wang[a,*], Meina Qiao[a], Yingjun Deng[b], Yi Zhou[c], Huan Wang[d,*], Qi Lyu[a], Hichem Snoussi[e]

[a] *School of Automation Science and Electrical Engineering, Beihang University, Beijing, China*
[b] *Center for Applied Mathematics, Tianjin University, Tianjin, China*
[c] *Department of Electronic Engineering, Dalian Maritime University, Dalian, China*
[d] *Department of Physics and State Key Laboratory of Low-Dimensional Quantum Physics, Tsinghua University, Beijing, China*
[e] *Institut Charles Delaunay-LM2S-UMR STMR 6281 CNRS, University of Technology of Troyes, France*

Abstract

Abnormal event detection is a challenging problem in video surveillance which is essential to the early-warning security and protection system. We propose an algorithm to solve this problem efficiently based on an image descriptor which encodes the movement information and the classification method. The new abnormality indicator is derived from the hidden Markov model which learns the histograms of optical flow orientations of the observed video frames. This indicator measures the similarity between the observed video frame and existing normal frames. The proposed method is evaluated and validated on several video surveillance datasets.
Keywords: abnormal event detection; optical flow; hidden Markov model
2016 MSC: 00-01, 99-00

1. Introduction

Video surveillance has become an important research area in computer vision. As a part of this subject, abnormal event detection is a key goal which

*Corresponding author
Email address: `wangtian@buaa.edu.cn` (Tian Wang), `wanghuan@phys.tsinghua.edu.cn` (Huan Wang)

has gained more and more attention. The point of abnormal event detection is to detect the behaviors that tend to be considered an anomalous or irregular pattern.

The abnormal event detection problem has been researched for several years. The anomalous behavior detection problem in discrete sequences was surveyed in [1], the proper models were given. In [2], different techniques for the fall detection which was in high demand within the security and the health-care industries were reviewed. In [3], the advances and trends in crowd analysis which was related to abnormal detection were surveyed in the context of crowd modeling studies. The contextual abnormal human behavior detection problem in video surveillance application was reviewed in [4]. The complexity of the video based on abnormal event detection was indicated. The influence of noisy data and representation of low-level features significantly influenced the discriminative power of the classifier. A systematic review of the technologies involved in anomaly detection in automated surveillance from front-end data acquisition to the information processing method was analyzed in [5], the presented technologies were categorized into surveillance target, anomaly definitions and assumptions, types of sensors used and the feature extraction processes, learning methods and modeling algorithms. In [6], an introduction to Variational Bayesian (VB) methods in the context of probabilistic graphical models was presented, and their application in multimedia related problems was discussed. In [7], the authors presented and evaluated human activity detection methodologies that utilized classification and statistical data mining methodologies. The movement representation and the event model algorithms are researched in these papers. In our work, the histogram of the optical flow orientations is adopted as the basic model as it describes the movement of the region of interest.

The probability models have been used for diverse applications in video analysis including abnormal detection. In [11], the behavior pattern was decomposed according to its temporal characteristics, and then modeled using the Dynamic Bayesian Networks. In [12], the crowd distribution information was represented by particle entropy and the GMM (Gaussian mixture model) over the normal crowd behaviors was used to predict the anomalies. In [13], a probabilistic Petri Net was proposed to recognize human activities in restricted settings such as airports, parking lots and banks, the minimal sub-videos in a given activity were identified with a probability above a certain threshold, and the activity from a given set with the highest probability was detected. In [14], a 2.5D graph, integrating 3D view-independent pose features and 2D appearance features, measured by matching methods was proposed for action image representation. In [15], a Bayesian network-based method was proposed for automatic event detection and summarization in soccer videos, and then seven different events in soccer videos were detected, namely, goal, card, goal attempt, corner, foul, offside and nonhighlights. While hidden Markov model (HMM) is a statistical analysis model that can process time sequential data, a global overview of the existing researches in abnormal event detection was introduced. In [8], HMM trained in a max-margin method was able to classify the transition and duration in highly varying videos. In [9], HMM was chosen as the underlying model to encode the activity concept transition in video events. Furthermore, video clips were treated as observations corresponding to latent activity concept variables in a HMM. In [10], a hierarchical dynamic framework firstly extracted high level skeletal joint features, and then the learned representation was recognized by deep belief networks which contain layers of features to predict probability distributions over states of HMM. In [16], multistream-fused HMM model was introduced to recognize the real-life visual behavior in a warehouse monitored by camera networks. In [17], the optical flow of the traffic area was extracted, and then HMM was used to detect abnormal events. In [18], HMM was used to identify uncommon motion events based on motion coding, which encoded the information of intrinsic dynamics. As the HMM has the powerful ability modeling the action, it is chosen as the prototypical method to analyze the video event in our work.

We propose an algorithm to detect abnormal events based on video analysis including the movement feature descriptor and the classification method.

(a) UMN scene (b) PETS scene

Figure 1: The scenarios in the UMN and PETS datasets. (a) A normal lawn scene in UMN dataset. (b) A normal scene (from Time14-55) in PETS dataset.

The feature descriptor encoding the movement information based on analyzing the optical flow of the region of interest is proposed. And then, the hidden Markov model is derived to distinguish hide states by analyzing the feature descriptor with probability property. The probability property of the histogram based feature descriptor is analyzed. Thus, the HMM model suits the abnormal classification application. The UMN [19] and PETS [20] are chosen as the benchmark datasets in this paper. The scenarios are shown in Fig. 1. UMN dataset simulates panic-driven scenes, and PETS dataset imitates suspicious moving queues.

The rest of the paper is organized as follows. In Section 2, the subject of the abnormal detection algorithm is proposed. Firstly, the histogram of optical flow orientations descriptor is briefly introduced. Afterwards, the hidden Markov models the descriptor, and then the classification method is proposed. In Section 3, experimental results of the benchmark datasets are illustrated and discussed. Finally, Section 4 concludes the paper.

2. Hidden Markov Model based on Histogram of Optical Flow Orientations

The abnormal detection method is proposed in this section. Firstly, the feature descriptor which computes the histogram of optical flow orientations (HOFO) is introduced. Owing to the probabilistic properties of the HOFO feature, the hidden Markov modeling (HMM) based classifier is proposed to distinguish the normal event from the abnormal event.

2.1. Histograms of Optical Flow Orientations

The feature descriptor, histogram of optical flow orientations (HOFO) for describing movement information from region of interest, is proposed in our previous work [21]. Firstly, the optical flow is extracted from consequential frames to obtain the low-level movement information. Horn-Schunck (HS) [22] algorithm is adopted. The computation over the optical flow field is shown in Fig. 2. The scene descriptor is computed over spatial blocks, with the optical flow orientation feature. Horizontal and vertical optical flow (u and v fields) are distributed into 9 orientation bins. While the histogram feature is closely related to the probability distribution, we are inspired to handle the abnormal detection problem via the probabilistic graphical models. By choosing n bands, the HOFO for a frame is given by a n-dimensional vector $\mathbf{q}$.

$$\mathbf{q} = (q_1, \cdots, q_n), \tag{1}$$

with $\sum_{i=1}^{n} q_i = 1$.

2.2. Hidden Markov Modeling

It is assumed that the HOFO feature descriptors of the normal and abnormal video frames exist in the training step. Based on learning the training samples, we propose the abnormal event detection method derived from HMM.

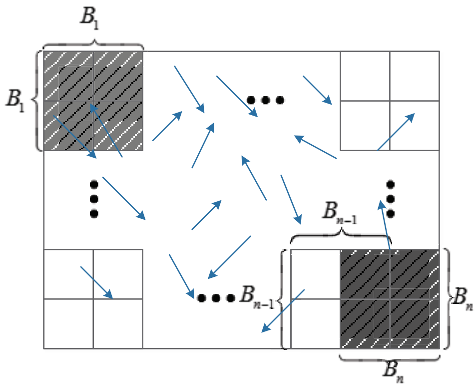


Figure 2: Computation of the HOFO feature descriptor on the image over the optical flow fields.

2.2.1. *General HMM*

As described in [23], for a general HMM with m hidden states and n observable states, the stationary distributions of the observable states are

$$(p_{i1}, \cdots, p_{in}),$$

where $i \in 1, 2, ..., m$ with $\sum_{j=1}^{n} p_{ij} = 1$, the stationary distribution of the hidden states is

$$(\alpha_1, \alpha_2, ..., \alpha_m),$$

with $\sum_{i=1}^{m} \alpha_i = 1$.

Then the transition probability matrix for all the observable states should be

$$
\begin{aligned}
P =& \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{m1} \\ p_{12} & p_{22} & \cdots & p_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1n} & p_{2n} & \cdots & p_{mn} \end{bmatrix} \begin{bmatrix} \alpha_1 & \alpha_1 & \cdots & \alpha_1 \\ \alpha_2 & \alpha_2 & \cdots & \alpha_2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_m & \alpha_m & \cdots & \alpha_m \end{bmatrix} \\
=& \mathbf{p}(1, 1, \cdots, 1),
\end{aligned} \tag{2}
$$

subject to

$$
\mathbf{p} = \begin{bmatrix} \sum_{k=1}^{m} \alpha_k p_{k_1} \\ \sum_{k=1}^{m} \alpha_k p_{k_2} \\ \vdots \\ \sum_{k=1}^{m} \alpha_k p_{k_m} \end{bmatrix}. \tag{3}
$$

It is noted that the vector $\mathbf{p}$ is the stationary probability distribution of P [23]. It is guessed that the distance between this stationary distribution and the observed distribution can be utilized for the classification problems. This idea motivates our consideration in this paper.

2.2.2. Normal and Abnormal Event Modeling

The HOFO feature of a video frame is perfect for applying the HMM due to its probabilistic property. As the HOFO is already a histogram of observed records, no extra statistical treatments are needed. For the abnormal event detection problem in video, there are $m = 2$ hidden states: $state1$ for normal frames and $state2$ for abnormal frames. The observable state n is the dimension of the feature descriptor. The dimension of the proposed HOFO descriptor is $n = 36$.

Suppose that there are K_1 normal video frames and K_2 abnormal frames of the training samples. This leads to n-dimensional HOFO series $\{\mathbf{q}_1^j\}_{j=1}^{K_1}$ and

$\{\mathbf{q}_2^j\}_{j=1}^{K_2}$ for normal and abnormal frames, respectively. Here the HOFO $\mathbf{q}_i^j$ is a n-dimensional vector:

$$\mathbf{q}_i^j = (q_{i1}^j, ..., q_{in}^j), \quad i = 1, 2; \; j = 1, \cdots, K_i. \tag{4}$$

Consider that the two average distributions $\mathbf{p}_1$ and $\mathbf{p}_2$ is based on observed HOFO series in normal and abnormal frames:

$$\mathbf{p}_i := (p_{i1}, \cdots, p_{in}) = \frac{1}{K_i} \sum_{j=1}^{K_i} \mathbf{q}_i^j, \quad i = 1, 2. \tag{5}$$

In a stable environment, the HOFO feature for a video frame should also be stable. Therefore, $\mathbf{p}_1$ and $\mathbf{p}_2$ can be treated as stationary distributions of observations under different hidden states. From a geometric view, $\mathbf{p}_i$ is the closest point to observed points $\{\mathbf{q}_i^j\}_{j=1}^{K_i}, i = 1, 2$. $\mathbf{p}_i, i = 1, 2$ is the solution to the following minimization problem

$$\min_{\mathbf{p} \in \mathbb{R}^n} \sum_{j=1}^{K_i} \|\mathbf{p} - \mathbf{q}_i^j\|_2^2. \tag{6}$$

The stationary probability is denoted for the normal frame as α, correspondingly $1-\alpha$ for the abnormal frame. For the video frame event detection problem, the transition matrix (7) is updated:

$$\begin{aligned} P = & \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \\ \vdots & \vdots \\ p_{1n} & p_{2n} \end{bmatrix} \begin{bmatrix} \alpha & \alpha & \cdots & \alpha \\ 1-\alpha & 1-\alpha & \cdots & 1-\alpha \end{bmatrix} \\ = & \mathbf{p}(1, 1, \cdots, 1), \end{aligned} \tag{7}$$

where

$$\mathbf{p} = \begin{bmatrix} \alpha p_{11} + (1-\alpha) p_{21} \\ \alpha p_{12} + (1-\alpha) p_{22} \\ \vdots \\ \alpha p_{1n} + (1-\alpha) p_{2n} \end{bmatrix} = \alpha \mathbf{p}_1 + (1-\alpha) \mathbf{p}_2. \tag{8}$$

$\mathbf{p}$ is the observed distribution for observable states. Then based on observed video frames, the unknown stationary probability α should be found by normal estimation methods, such as maximum likelihood estimation and expectation-maximization algorithm. In this paper instead of estimating the HMM itself, a two-class classification method will be introduced in the next subsection.

2.3. Anomaly Detection from Two-Class Classification

In the test step, the HOFO descriptor of each frame can be calculated and the observed distribution $\mathbf{q}$ of the observable state can be obtained. The stationary distribution α of hidden states can be estimated by maximizing the difference between $\mathbf{q}$ and the predicted distribution $\mathbf{p}$. In another word, α needs to be solved from the following minimization problem [23]

$$\min_{\alpha} \|\mathbf{p} - \mathbf{q}\|_2^2, \tag{9}$$

subject to

$$0 \leq \alpha \leq 1. \tag{10}$$

Substituting (8) into (9), we have

$$\begin{aligned} & \min_{0\leq\alpha\leq 1} \|\mathbf{p} - \mathbf{q}\|_2^2 \\ = & \min_{0\leq\alpha\leq 1} \|\alpha\mathbf{p}_1 + (1-\alpha)\mathbf{p}_2 - \mathbf{q}\|_2^2 \\ = & \min_{0\leq\alpha\leq 1} \left\{\alpha^2 \|\mathbf{p}_1 - \mathbf{p}_2\|_2^2 - 2\alpha < \mathbf{q} - \mathbf{p}_2, \mathbf{p}_1 - \mathbf{p}_2 > + \|\mathbf{q} - \mathbf{p}_2\|_2^2\right\}. \end{aligned} \tag{11}$$

As the objective function in (11) is a quadratic function of α, we consider its axis of symmetry,

$$\tau(\mathbf{q}) = \frac{< \mathbf{q} - \mathbf{p}_2, \mathbf{p}_1 - \mathbf{p}_2 >}{< \mathbf{p}_1 - \mathbf{p}_2, \mathbf{p}_1 - \mathbf{p}_2 >}. \tag{12}$$

Then the solution to the optimization problem (11) is given by

$$\alpha(\mathbf{q}) = \begin{cases} 0 & \text{if} \quad \tau(\mathbf{q}) \leq 0; \\ \tau(\mathbf{q}) & \text{if} \quad 0 < \tau(\mathbf{q}) < 1; \\ 1 & \text{if} \quad \tau(\mathbf{q}) \geq 1. \end{cases} \tag{13}$$

The indicator (13) is found by a similarity measure for the observed $\mathbf{q}$ to the normal indicator $\mathbf{p}_1$ and abnormal indicator $\mathbf{p}_2$. From the geometric view, the smaller $\alpha(\mathbf{q})$ is, the closer $\mathbf{q}$ is to $\mathbf{p}_1$.

$\alpha(\mathbf{q})$ is an anomaly indicator which is irrelevant to the HMM itself. And it can be used as an independent indicator like other similarity measures without referring to the knowledge of HMM. Actually, when an alarm threshold l is set, a video frame with HOFO feature $\mathbf{q}$ is then easily classified:

$$\begin{cases} \text{normal frame: } \alpha(\mathbf{q}) > l; \\ \text{abnormal frame: } \alpha(\mathbf{q}) \leq l. \end{cases} \tag{14}$$

The alarm level l is set from the minimization problem for the error function based on historical video frames:

$$\min_{0 \leq l \leq 1} \left[\sum_{j=1}^{K_1} I(\alpha(\mathbf{q}_1^j) < l) + \sum_{j=1}^{K_2} I(\alpha(\mathbf{q}_2^j) \geq l) \right]. \tag{15}$$

To summarize all above, the algorithm for abnormal event recognition using HMM is given in Algorithm 1.

3. Abnormal event detection results

This section presents experiments of the HMM classification for abnormal event detection. UMN [19] and PETS [20] datasets are adopted to show the results.

3.1. UMN dataset

The UMN dataset has the lawn, indoor and plaza scenes of crowded escaping events. The normal sample is defined as the individual walking. The abnormal

Algorithm 1 HMM Classifier for Abnormal Event Detection

Require:

1. K_1 training normal frames, with HOFO features $\{\mathbf{q}_1^j\}_{j=1}^{K_1}$.
2. K_2 training abnormal frames, with HOFO features $\{\mathbf{q}_2^j\}_{j=1}^{K_2}$.
3. test frame with HOFO feature $\mathbf{q}$.

1: Calculate the normal and abnormal indicators $\mathbf{p}_1$ and $\mathbf{p}_2$ from (5).
2: Set the alarm level l from the minimization problem (15).
3: Abnormality classification for $\mathbf{q}$ from (13):

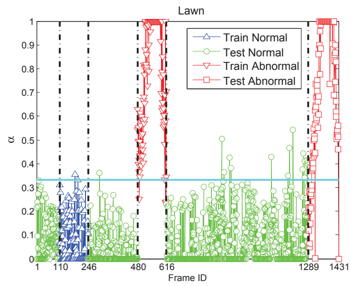
$$\begin{cases} \text{normal}: & \alpha(\mathbf{q}) > l; \\ \text{abnormal}: & \alpha(\mathbf{q}) \leq l. \end{cases}$$

sample is defined as the individual running in different directions, which implies that the people are escaping from some dangerous place. The detection results of the lawn scene are shown in Fig. 3. The normal and abnormal samples are chosen from the lawn scene itself. The results show that the histogram descriptor with the HMM classification method can distinguish the normal event from the abnormal one. From Fig. 3, we can see that the α value of some normal frames larger than the threshold, and then these frames are classified as abnormal ones. However, these false positive samples appear in an abrupt model. In other words, these isolated samples are in the state without any anterior or posterior consecutive ones to constitute a continuous state. Thus, the prediction result of these frames can be considered as the noise to a certain extent, and then be modified into the normal state. The accuracy rate remarked in the figure is counted based on the raw prediction result. But if the result is modified after the noise has been filtered, the accurate rate of the prediction can be increased.

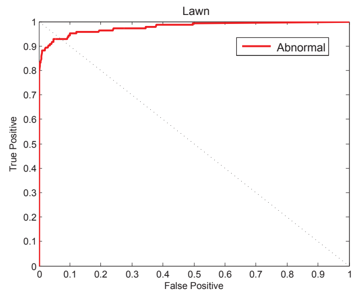
The detection results of the indoor scene are shown in Fig. 4. The accuracy rate is a little low than the one in lawn scene, and the false negative frames are at the beginning or the final part of each abnormal sequence. In the beginning

(a) Normal lawn scene

(b) Abnormal lawn scene



(c) Lawn scene results



(d) Lawn ROC

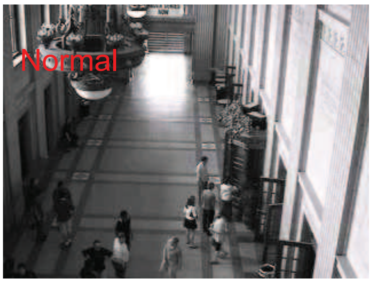
Figure 3: Lawn scene detection results. People walking is regarded as the normal event, whereas people running is regarded as the abnormal event. (a) One normal frame detected by the proposed method. (b) One abnormal frame detected by the proposed method. (c) The detection results of the lawn sequence. The *blue triangle*, *red triangle*, *green circle* and *red square* represent the training normal frame, training abnormal frame, testing normal frame and testing abnormal frame, respectively. The critical frame numbers are labeled. The detection accuracy is 97.24%. (d)The receiver operating characteristic (ROC) curve of the classification strategy for the lawn scene, the area under the ROC curve (AUC) of the detection results is 0.9779.

several frames of the abnormal sequence, the frame is in an ambiguous state. Furthermore, these frames can be classified as either normal or abnormal. In the frames at the end of the abnormal sequence, there are no people in the frame. Because the HOFO feature descriptor is based on the movement state of the frame, the index of the HOFO bin is in a low level. Thus, these frames contribute to the false negative, and then lead to the lower performance of the classifier. Whereas, these frames can be classified into the abnormity by detecting the people in the frame. In another word, the frames can be classified as abnormal directly if there does not exist any person within it. The illumination varies in several frames. Besides, the size of the multiple persons is different from the lawn scene and the plaza scene. However, the proposed abnormal event detection method can deal with the illumination and scale variation.

The detection result of the plaza scene is shown in Fig. 5. The false negative frame is at the beginning and the final part of each sequence. The frame with the wrong label is independent as well, in a similar way to the lawn scene. These frames can be filtered out. In Fig. 5, α values of the training frames from 380 to 410 are larger than the threshold. The representative frames in this sequence which are contrasted with the ones in the sequence from 1 to 380 are presented in Fig. 6. In Fig. 6(a), people are moving compactly in the center of the surveillance visual field. By comparing 6(b) with (a), people in other training frames are walking dispersedly, and are filling the whole frame. Consequently, these frames are treated as the outlier in the training samples. On the other hand, the HMM classification framework with the HOFO feature descriptor can not only distinguish the normal or abnormal state from the video, but also mines much more deep information of the samples.

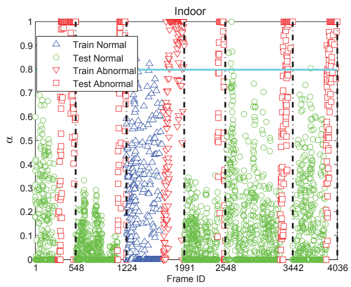
3.2. PETS dataset

PETS [20] benchmark dataset includes multisensor sequences containing different crowd activists. The aim of this dataset is to evaluate the algorithms or systems for the detection of crowd surveillance characteristics and events within a real-world environment. For example, the crowd count, detection of separate

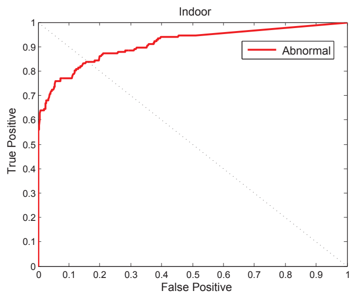


(a) Normal indoor scene

(b) Abnormal indoor scene



(c) Indoor scene results

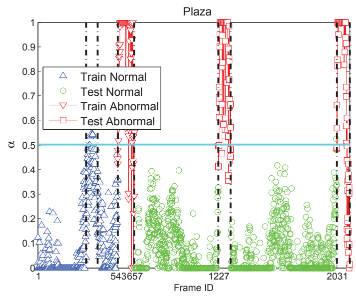


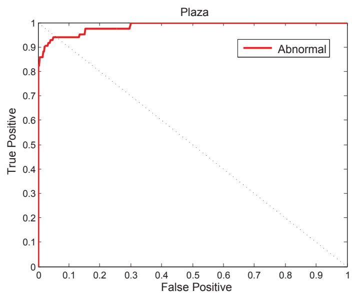
(d) Indoor ROC

Figure 4: Indoor scene detection results. (a) One normal frame detected by the proposed method. (b) One abnormal frame detected by the proposed method. (c) The detection results of the indoor sequence. The *blue triangle*, *red triangle*, *green circle* and *red square* represent the training normal frames, training abnormal frames, testing normal frames and testing abnormal frames, respectively. The critical frame numbers are labeled. The detection accuracy is 91.39%. (d)The ROC of the classification strategy for the indoor scene, the AUC is 0.9223.

(a) Normal plaza scene

(b) Abnormal plaza scene





(c) Plaza scene results

(d) Plaza ROC

Figure 5: Plaza scene detection results. (a) One normal frame detected by the proposed method. (b) One abnormal frame detected by the proposed method. (c) The detection result of the plaza sequence. The *blue triangle*, *red triangle*, *green circle* and *red square* represent the training normal frames, training abnormal frames, testing normal frames and testing abnormal frames, respectively. The critical frame numbers are labeled. The detection accuracy is 97.27%. (d)The ROC of the classification strategy for the plaza scene, the AUC is 0.99849.

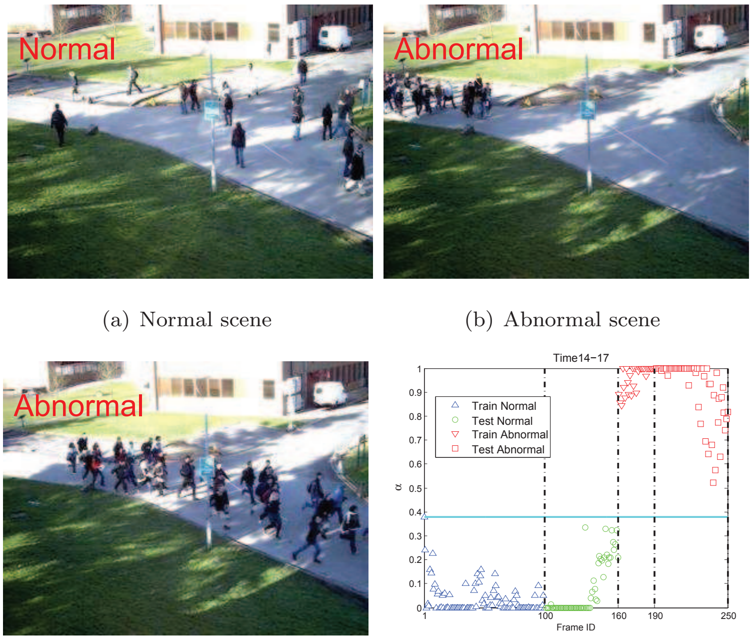
(a) Normal plaza scene

(b) Abnormal plaza scene

Figure 6: Training frames of plaza scene. (a) One normal training frame. (b) One training frame which is treated as the outlier by the proposed HMM based classification method.

Method	Area under ROC		
	lawn	indoor	plaza
Social Force [24]	0.96		
Optical Flow [24]	0.84		
NN [25]	0.93		
SRC [25]	0.995	0.975	0.964
STCOG [26]	0.9362	0.7759	0.9661
OC-SVM [21]	0.9845	0.9037	0.9815
HMM with optical flow (ours)	0.9779	0.9223	0.9849

Table 1: The performance of our proposed HMM with optical flow based feature method compared with other state-of-the-art methods.



(a) Normal scene (b) Abnormal scene

(c) Abnormal scene (d) Results

Figure 7: Time14-17 scene detection results. (a) One normal frame detected by the proposed method. (b,c) Abnormal frames detected by the proposed method. (d) The detection results of the sequence. The critical frame numbers are labeled. The detection accuracy is 100%, and the AUC is 1.

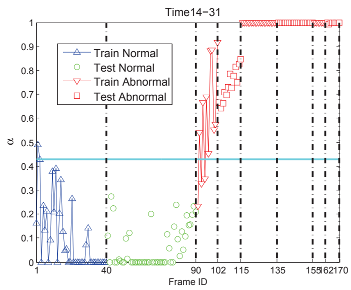
flows and specific crowd events are in the research area of the PETS dataset. The specific crowd event detection problem is studied in our work in the HOFO with HMM classification framework. Two scenes Time14-17 and Time14-31 are analyzed in this paper. The experiments differentiate the splitting queues from the one queue, and distinguish the people moving in the same direction from the loitering.

The detection results of the Time14-17 scene are shown in Fig. 7. Normal samples are chosen from the Time14-55 scene, while abnormal samples are chosen from the Time14-17 scene. In the normal scene, people are walking in different directions. Correspondingly, people moving in the same direction is considered as the abnormal event, which implies some parade or assembly. With the proposed classification method, the abnormal event is detected with high accuracy.

(a) Cohesive crowd

(b) Splitting crowd

(c) Splitting crowd



(d) Results

Figure 8: Time14-31 scene detection results. The cohesive crowd is regarded as the normal event, whereas the splitting crowd is detected as the abnormal event. (a) One cohesive crowd scene detected by the proposed method. (b,c) Splitting crowd scenes detected by the proposed method. (d) The detection results of the sequence. The critical frame numbers are labeled. The detection accuracy is 100%, and The AUC is 1.

The detection results of the Time14-31 scene are shown in Fig. 8. Normal training samples are chosen from the Time14-16 sequence, and the persons are walking in one queue. Abnormal samples and the normal testing samples are chosen from Time14-31 itself. The splitting crowds, considered as normal scenes, are distinguished from the cohesive ones by the proposed algorithm based on the HOFO feature and HMM with high accuracy.

4. Conclusions

An abnormal event detection method is proposed in this paper. The histogram of optical flow orientation descriptor (HOFO) is used to extract the frame movement information. Owing to the probability property of the HOFO feature descriptor, a classification method is derived from the hidden Markov model to distinguish the abnormal event from the normal event. Our experimental studies demonstrate the advantage of the proposed method. In the future work, the multi-class problem will be tested to show the robustness of the proposed algorithm.

Acknowledgment

This work is partially supported by the National Natural Science Foundation of China (Grant No. 61503017, U1435220, 71701143), the China Postdoctoral Science Foundation (Grant No. 2016M591156), the Fundamental Research Funds for the Central Universities (Grant No. YWF-14-RSC-102), the ANR AutoFerm project and the Platform CAPSEC funded by Région Champagne-Ardenne and FEDER.

References

[1] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection for discrete sequences: A survey, IEEE Transactions on Knowledge and Data Engineering 24 (5) (2012) 823–839.

[2] M. Mubashir, L. Shao, L. Seed, A survey on fall detection: Principles and approaches, Neurocomputing 100 (2013) 144–152.

[3] M. S. Zitouni, H. Bhaskar, J. Dias, M. Al-Mualla, Advances and trends in visual crowd analysis: A systematic survey and evaluation of crowd modelling techniques, Neurocomputing 186 (2016) 139–159.

[4] O. P. Popoola, K. Wang, Video-based abnormal human behavior recognition—a review, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 42 (6) (2012) 865–878.

[5] A. Sodemann, M. P. Ross, B. J. Borghetti, et al., A review of anomaly detection in automated surveillance, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 42 (6) (2012) 1257–1272.

[6] Z. Chen, S. Derin Babacan, R. Molina, A. K. Katsaggelos, Variational bayesian methods for multimedia problems, IEEE Transactions on Multimedia 16 (4) (2014) 1000–1017.

[7] S. K. Tasoulis, C. N. Doukas, V. P. Plagianakos, I. Maglogiannis, Statistical data mining of streaming motion data for activity and fall recognition in assistive environments, Neurocomputing 107 (2013) 87–96.

[8] K. Tang, L. Fei-Fei, D. Koller, Learning latent temporal structure for complex event detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1250–1257.

[9] C. Sun, R. Nevatia, Active: Activity concept transitions in video event classification, in: Proceedings of IEEE International Conference on Computer Vision (ICCV), 2013, pp. 913–920.

[10] D. Wu, L. Shao, Leveraging hierarchical parametric networks for skeletal joints based action segmentation and recognition, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2014, pp. 724–731.

[11] C. C. Loy, T. Xiang, S. Gong, Detecting and discriminating behavioural anomalies, Pattern Recognition 44 (1) (2011) 117–132.

[12] X. Gu, J. Cui, Q. Zhu, Abnormal crowd behavior detection by using the particle entropy, Optik-International Journal for Light and Electron Optics 125 (14) (2014) 3428–3433.

[13] M. Albanese, R. Chellappa, V. Moscato, A. Picariello, V. Subrahmanian, P. Turaga, O. Udrea, A constrained probabilistic petri net framework for human activity detection in video, IEEE Transactions on Multimedia 10 (6) (2008) 982–996.

[14] B. Yao, L. Fei-Fei, Action recognition with exemplar based 2.5 d graph matching, in: European Conference on Computer Vision (ECCV), Springer, 2012, pp. 173–186.

[15] M. Tavassolipour, M. Karimian, S. Kasaei, Event detection and summarization in soccer videos using bayesian network and copula, IEEE Transactions on Circuits and Systems for Video Technology 24 (2) (2014) 291–304.

[16] D. Kosmopoulos, S. P. Chatzis, Robust visual behavior recognition, IEEE Signal Processing Magazine 27 (5) (2010) 34–45.

[17] Á. Utasi, L. Czúni, Detection of unusual optical flow patterns by multilevel hidden markov models, Optical Engineering 49 (1) (2010) 017201–017201.

[18] H. Jiménez-Hernández, J.-J. González-Barbosa, T. Garcia-Ramírez, Detecting abnormal vehicular dynamics at intersections based on an unsupervised learning approach and a stochastic model, Sensors 10 (8) (2010) 7576–7601.

[19] UMN, Unusual crowd activity dataset of university of minnesota, department of computer science and engineering, http://mha.cs.umn.edu/movies/crowd-activity-all.avi.

[20] PETS, Performance evaluation of tracking and surveillance (pets) 2009 benchmark data. multisensor sequences containing different crowd activities. http://www.cvg.rdg.ac.uk/pets2009/a.html.

[21] T. Wang, H. Snoussi, Detection of abnormal visual events via global optical flow orientation histogram, IEEE Transactions on Information Forensics and Security 9 (6) (2014) 988–998.

[22] B. K. Horn, B. G. Schunck, Determining optical flow, Artificial intelligence 17 (1) (1981) 185–203.

[23] W. Ching, M. K. Ng, Markov Chains: Models, Algorithms and Applications, Springer, 2006.

[24] R. Mehran, A. Oyama, M. Shah, Abnormal crowd behavior detection using social force model, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami,FL,USA, June, 2009, pp. 935–942.

[25] Y. Cong, J. Yuan, J. Liu, Sparse reconstruction cost for abnormal event detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June, 2011, pp. 3449–3456.

[26] Y. Shi, Y. Gao, R. Wang, Real-time abnormal event detection in complicated scenes, in: Proceedings of International Conference on Pattern Recognition (ICPR), Istanbul, Turkey, August, 2010, pp. 3653–3656.