REAL-TIME ANOMALY DETECTION WITH HMOF FEATURE

Huihui Zhu, Bin Liu, Guojun Yin, Yan Lu, Weihai Li, Nenghai Yu

CAS Key Laboratory of Electromagnetic Space Information
University of Science and Technology of China, Hefei, China

ABSTRACT

Anomaly detection is a challenging problem in intelligent video surveillance. Most existing methods are computation-consuming, which cannot satisfy the real-time requirement. In this paper, we propose a real-time anomaly detection framework with low computational complexity and high efficiency. A new feature, named Histogram of Magnitude Optical Flow (HMOF), is proposed to capture the motion of video patches. Compared with existing feature descriptors, HMOF is more sensitive to motion magnitude and more efficient to distinguish anomaly information. The HMOF features are computed for foreground patches, and are reconstructed by the auto-encoder for better clustering. Then, we use Gaussian Mixture Model (GMM) Classifiers to distinguish anomalies from normal activities in videos. Experimental results show that our framework outperforms state-of-the-art methods, and can reliably detect anomalies in real-time.

Index Terms— Anomaly detection, HMOF, auto-encoder, real-time

1. INTRODUCTION

Anomaly detection and localization in intelligent video surveillance is a significant task due to the growing needs of public security. In real life, the definition of abnormalities in the video is varied. For example, a runner is seen as normal on the track and field, while it will be regarded as abnormal in the square. Therefore, it is difficult for us to use the same standard to measure all the scenes. A video event is usually considered as an anomaly if it is not very likely to occur in the video [1]. Thus we need the normal monitor video of the scene to establish a normal model, which identifies the anomaly in the detection.

In recent studies, the sparse representations of events [2] in videos have been widely explored. The proposed models in [2, 3] achieve favorable performance in global abnormal events (GAE), however they often fail in the local abnormal events (LAE). In order to solve this problem, some methods have proposed trajectory-based anomaly detection methods, such as particle trajectories [4], tracking trajectories [5] and so on. Such methods tend to behave very well in simple sparse scenes, but their performance in complex scenes is severely degraded. Some methods take account into the energy distribution characteristics of the crowd. When anomaly exists, some energy can be mutated, such as pedestrian loss model [6]. Some methods divide frames of video into numbers of patches, and then can detect anomalies in LAE by analyzing the patches [7, 8]. In addition, some methods utilize features based on the optical flow, such as Histograms of Oriented Optical Flow (HOF) [9] and Multi-scale Histogram of Optical Flow (MHOF) [1]. These algorithms are time-consuming and can hardly meet the real-time requirements.

Sabokrou et al [10] use two feature descriptors to extract global and local features. Roberto et al [11] extract the HOF features of the foreground area to build a dictionary to detect anomalies. These algorithms can meet the real-time requirements with high detection speeds. However, compared with state-of-the-art methods, there are still some gaps in detection performance.

In this paper, we propose a new feature called Histogram of Magnitude Optical Flow (HMOF), which is more efficient to describe motion information. We use the foreground extraction algorithm to extract the video foreground patches, so that only the foreground patches will be processed, which is more efficient. Next, the features are fed into the auto-encoder network to be reconstructed and then classified by the Gaussian Mixture Model (GMM) Classifiers. Thus we can distinguish the abnormal patches.

The main contributions of our work are as follows: (1) We present a new feature named HMOF for anomaly detection. Compared with existing feature descriptors, HMOF is more sensitive to motion magnitude and more efficient to distinguish anomaly information; (2) We propose an algorithm framework for local anomaly detection, which can be applied in real scenes. Because our algorithm outperforms state-of-the-art methods, and can be done in real-time.

The rest of the paper is organized as follows. The proposed method is introduced in Section 2. Section 3 presents the experimental results, comparisons and analysis on UMN and UCSD datasets. Finally, Section 4 concludes the work.
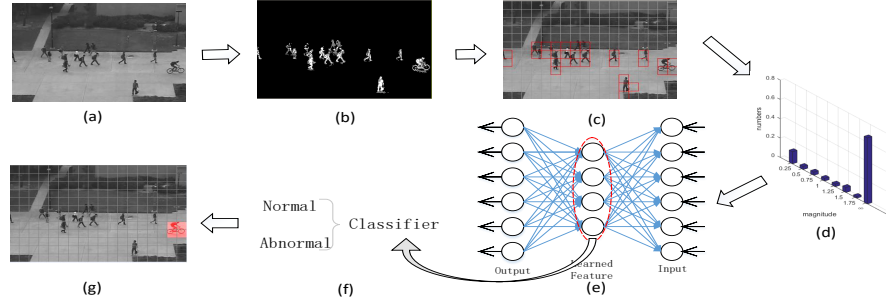
Fig. 1. Framework of our proposed method. (a) Input frames. (b) Extracted foreground. (c) Foreground patches. (d) HMOF features. (e) Auto-Encoder. (f) GMM Classifier. (g) Detected anomalies.

2. THE PROPOSED METHOD

In this section, we illustrate the proposed algorithm in detail. Firstly, we obtain foregrounds patches with KNN matting. Secondly, HMOF features are extracted on foreground patches. Base on the features, we use the auto-encoder network to get the deep features, which will be fed into the GMM Classifiers. The Framework of our method is shown in Fig.1.

2.1. Foreground detection

We split each frame into a number of non-overlapping patches. In order to reduce the number of processing patches, we use foreground detection algorithm to extract foreground area. As a matter of fact, background patches are eliminated, which can speeds up the test phase.

The problem of foreground segmentation is treated as matting, and the matting model is expressed as follows:

$$I = aF + (1 - a)B \qquad (1)$$

where I, F, B is color, foreground color and background color of a pixel in image respectively, a is a parameter of segmentation to indicate the former background weight. Here we use the well-known KNN matting algorithm [12] to extract the foreground. The extracted foreground is shown in Fig.1 (b). Then we calculate the foreground value of each patch by adding the intensity of every pixel. If the value exceeds a threshold, we regard it as a foreground patch. The extracted foreground patches are shown in Fig.1 (c).

2.2. HMOF

Optical flow is good for describing the motion, and HOF is widely used as a motion descriptor. To extract HOF features, the amplitude weighting statistics of the optical flow is calculated in different directions of the optical flow, and the histogram of the optical flow direction information is obtained. However, the HOF features mainly consider the information of the optical flow direction, with less consideration of the

optical flow amplitude information. MHOF [1] is based on the HOF, taking account into the optical flow amplitude information, by setting the relevant threshold information for different amplitude range of different optical flow to carry out statistics.

While MHOF uses the amplitude characteristics of the optical flow, it still mainly considers the direction of the optical flow information, which means that MHOF is less sensitive to the motion magnitude. Furthermore, the amplitude threshold is usually an experience parameter. Generally, abnormal behaviors are more sensitive to the amplitude characteristics rather than directional characteristics of the optical flow, such as running, bicycles, cars, skate, etc. The speed of these behaviors are faster than the speed of normal behaviors. To some extent, the directional characteristics of the optical flow is a kind of interference. To reach a better performance, we propose a new motion feature called HMOF based on the amplitude characteristics of the optical flow, which can detect abnormal objects effectively.

The procedure of HMOF feature extraction is shown in Fig.2. Firstly, we need to calculate the threshold δ of HMOF. We sort the amplitudes of optical flow in normal patches of the whole training set in ascending order. Since there inevitably exists some noise when calculating optical flow, we discard the top *5%* of the optical flow and set δ as the maximum amplitude of the remaining optical flow. Then we divide the amplitude of the optical flow into *n* bins. The range of *i*-th bin is *[(i-1)/n×δ, i/n×δ)*. In order to accommodate all the optical flow during the test phase, the range of the last bin is set to *[(n-1)/n×δ, +∞)*. After that, we use the normalized histogram to keep the scale invariance of the HMOF feature.

Fig.3 shows the feature maps of HOF, MHOF and HMOF, from left to right respectively. Among them, the pedestrian and the tree are normal, and the bicycle and the car are regarded as abnormal. It can be seen that the HMOF is more prominent than the HOF and MHOF, and the characteristic distribution is more obvious. The feature distribution of the normal region is more biased towards the low amplitude side, while the abnormal area characteristic distribution is more
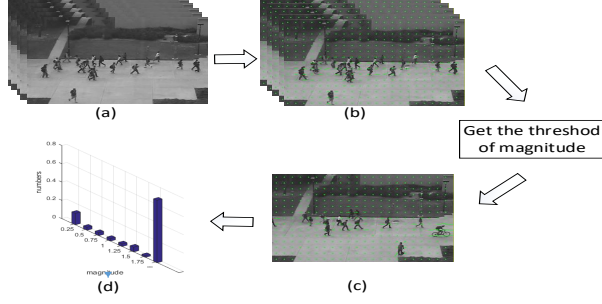
Fig. 2. The process of extracting HMOF features. (a) Videos of the training set. (b) Optical flow of the training set. (c) An area to be detected. (d) HMOF of the area.

biased towards the high side, which is conducive to distinguish between abnormalities. Subsequent experiments show that HMOF features perform better than the HOF and MHOF features.

2.3. Auto-encoder

Auto-encoder, consisting of encoder and decoder, is widely used in computer vision area such as video tracking and anomaly detection. Encoder aims to project the input data into the feature space which is constructed by the hidden units, then the input is reconstructed by the decoder. Auto-encoder is usually trained by minimizing the reconstruction error between the input and output. For supervised machine learning, the hidden layer of auto-encoders can be used for feature transformation.

In our case, the space expanded by the hidden units is called the feature space H. In the training phase, we use HMOF features of the training set to train the parameters of the auto-encoder. In the testing phase, we project the HMOF features of input data into feature space H. It can be seen that the features distribution of the normal and abnormal samples are quite different in the feature space H, which can be easily distinguished by the subsequent classifier.
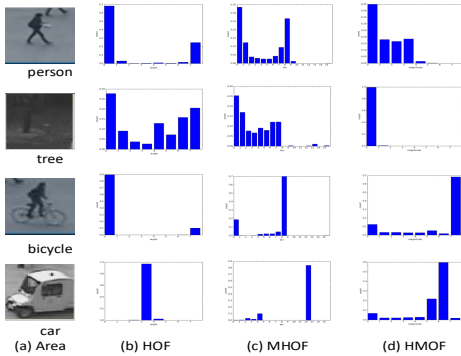

Fig. 3. the feature maps of HOF, MHOF and HMOF

2.4. Anomaly Classifier

The GMM is a weighed sum of multivariate Gaussian probability densities given by:

$$P(\mathbf{x}|\Theta) = \sum_{k=1}^{K} \lambda_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \qquad (2)$$

where $\Theta = \{\lambda_1, \cdots, \lambda_K, \boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_k\}$ is the parameter of GMM. K denote the number of Gaussian components and λ_k is the weight of the k-th Gaussian model. $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the mean and covariance matrix respectively. $\mathcal{N}(\cdot)$ denotes the multivariate Gaussian distribution. The parameters can be estimated by using the maximum-likelihood (ML) estimation. With the GMM method, we can adaptively adjust the decision surface for classification, which can better distinguish anomalies from normal activities in videos.

At first we use all features of the training set to train the GMM Classifiers, then the trained classifier is used to test the features of the testing set. Each feature will get a score after passing the classifier. If it is below a threshold α, it will be judged as an abnormal:

$$Patch(\mathbf{x}) = \begin{cases} Normal & score(\mathbf{x}) > \alpha \\ Abnormal & otherwise \end{cases} \qquad (3)$$

where $\mathbf{x}$ is the feature fed into the classifiers, and $score(\cdot)$ is the score given by the GMM Classifiers.

In some surveillance scenes, a target may be the formation of several patches due to the limitation of patch-based method. Thus in order to enhance the robustness, we judge a frame as abnormal if the number of abnormal patches exceeds a threshold β. If the number of abnormal patches is below the β, these patches are more likely to be misjudged and we will drop them out from the abnormal candidates:

$$Frame(i) = \begin{cases} Normal & sum(\mathrm{p}) < \beta \\ Abnormal & otherwise \end{cases} \qquad (4)$$

where p are the abnormal patches, $sum(\cdot)$ is the number of the abnormal patches in the i-th frame.

3. EXPERIMENTS

We use two measures to evaluate the results: the frame-level and the pix-level. For the frame-level, if one pixel is detected as an anomaly, the whole frame is considered as an anomaly. For the pixel-level, a frame is deemed to be correctly classified if at least 40% of the pixels are correctly classified [13]. We test our algorithm on two datasets: the UMN dataset for GAE detection and the UCSD dataset for LAE detection. The details are shown below.
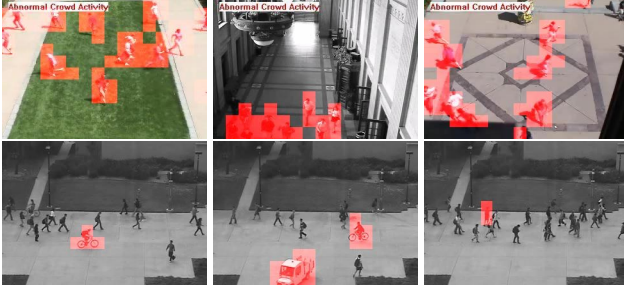
3.1. Detection of GAE on UMN dataset

The UMN dataset has three different scenes with a resolution of 320×240. In each scene, a group of people are walking

Table 3: Details of running-time (second/frame) on UCSD Ped2

Time (spf)	Foreground	Optical Flow	HMOF	Auto-Encoder	GMM	Total
Ours method	0.011	0.025	0.004	0.006	0.002	0.048

Table 1: Comparison of AUC on the UMN dataset

Scene	SR[1]	Zh[14]	MIP[15]	Scan[16]	Ours
1	99.5%	99.3%	99.6%	99.1%	**99.8%**
2	97.5%	96.9%	94.4%	95.1%	**98.6%**
3	96.4%	98.8%	90.8%	99%	**99.2%**

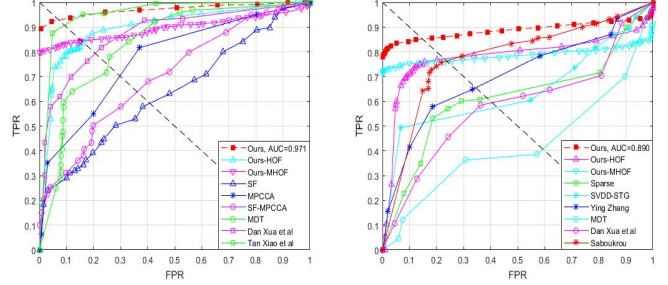**Fig. 4**. Examples of anomaly detection on UMN (top row) and UCSD Ped2 (bottom row)

in an area, and suddenly all people run away, which is considered to be abnormal. This dataset has no pixel-level ground truth, so we use Area Under the Curve (AUC) of the frame-level to evaluate our method.

We set the patch at a size of 20×20, and the amplitude of the optical flow is divided into 8 bins. The threshold δ of HMOF is calculated as 1.04 and the β used to adjudge abnormal frame is set to 3. Some image results are shown in Fig.4. We compare our method with SR [1], Zh [14], MIP [15], Scan [16]. Results are shown in Table 2, which demonstrate that our method outperforms state-of-the-art methods.

3.2. Detection of LAE on UCSD Ped2 dataset

The UCSD Ped2 dataset has 16 training and 12 testing video clips, and the number of frames of each clip varies. The videos consist of walking pedestrians paralleling to the camera plane, which are recorded with a static camera at 10 fps.

We set the patch at a size of 20×20, and the amplitude of the optical flow is divided into 8 bins. The threshold δ of HMOF is calculated as 2.4 and the β is set to 3. Some image results are shown in Fig.4. Our algorithm can detect bikers, cars, skaters, etc. Fig.5 shows the frame-level and pixel-level Receiver Operating Characteristic (ROC) of the UCSD Ped2. Equal Error Rate (EER) for the frame-level and the pixel-level comparisons is shown in Table 2. From Table 2, we can see that if the HMOF features are replaced by HOF or MHOF in the proposed method, the performance is much worse, which indicates the effectiveness of the HMOF features. We can also

Fig. 5. ROC comparison with state-of-the-art methods. Left: Frame-level. Right: Pixel-level

Table 2: EER for frame-level (FL) and pixel-level (PL) comparisons on UCSD Ped2; we only list first author in this table)

Method	FL	PL	Method	FL	PL
MDT [13]	24%	54%	Saligrama [7]	18%	-
Reddy [8]	21%	31%	Ying Zhang [17]	22%	33%
Dan [9]	20%	42%	Sabokrou [10]	19%	24%
Rosh [18]	17%	30%	DeepCascade [19]	8.2%	19%
Li [3]	18.5%	29.9%	Tan Xiao [20]	10%	17%
IBC [21]	13%	26%	Ours-MHOF	15.5%	23.9%
Ours-HOF	16.4%	22.8%	Ours	**7.2%**	**14.8%**

see from Table 2, compared with other state-of-the-art methods, the proposed algorithm achieves the best performance, with the frame-level EER decreased from 8.2% to 7.2%, and the pixel-level EER decreased from 17% to 14.8%.

3.3. Running-Time Analysis

The experiments are conducted on a regular PC with Intel-i7-7700 CPU (3.6 GHz) and 8 GB RAM, and the running-time of processing a single frame of UCSD Ped2 is provided in Table 3. Our method is computational efficient with the total time for detecting an anomaly in a frame being 0.048sec, which indicates that our method can be proceed in real-time.

4. CONCLUSION

In this paper, we present an anomaly detection method. A new feature named HMOF is proposed. Compared with other feature descriptors, HMOF is more sensitive to motion magnitude, and efficient to represent anomaly information. In our method, HMOF is computed for each foreground area, and is reconstructed by the auto-encoder. Then we use GMM Classifiers to distinguish anomalies from normal activities in videos. Experimental results show that our algorithm outperforms state-of-the-art methods, and can reliably detect

anomalies in real-time. Therefore, it can be widely used in real-time surveillance applications.

5. REFERENCES

[1] Yang Cong, Junsong Yuan, and Ji Liu, "Sparse reconstruction cost for abnormal event detection," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 3449–3456.

[2] Yang Cong, Junsong Yuan, and Yandong Tang, "Video anomaly search in crowded scenes via spatio-temporal motion context," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 10, pp. 1590–1599, 2013.

[3] Weixin Li, Vijay Mahadevan, and Nuno Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 1, pp. 18–32, 2014.

[4] Shandong Wu, Brian E Moore, and Mubarak Shah, "Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2054–2060.

[5] Zhang Jing, Gao Wei, Liu Anan, Gao Zan, Su Yuting, and Zhang Zhe, "Modeling approach of the video semantic events based on motion trajectories," *Electronic Measurement Technology*, vol. 9, pp. 008, 2013.

[6] Paul Scovanner and Marshall F Tappen, "Learning pedestrian dynamics from the real world," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 381–388.

[7] Venkatesh Saligrama and Zhu Chen, "Video anomaly detection based on local statistical aggregates," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2112–2119.

[8] Vikas Reddy, Conrad Sanderson, and Brian C Lovell, "Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*. IEEE, 2011, pp. 55–61.

[9] Dan Xu, Rui Song, Xinyu Wu, Nannan Li, Wei Feng, and Huihuan Qian, "Video anomaly detection based on a hierarchical activity discovery within spatio-temporal contexts," *Neurocomputing*, vol. 143, pp. 144–152, 2014.

[10] Mohammad Sabokrou, Mahmood Fathy, Mojtaba Hoseini, and Reinhard Klette, "Real-time anomaly detection and localization in crowded scenes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 56–62.

[11] Roberto Leyva, Victor Sanchez, and Chang-Tsun Li, "Video anomaly detection with compact feature sets for online performance," *IEEE Transactions on Image Processing*, 2017.

[12] Qifeng Chen, Dingzeyu Li, and Chi-Keung Tang, "Knn matting," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 9, pp. 2175–2188, 2013.

[13] Vijay Mahadevan, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos, "Anomaly detection in crowded scenes," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 1975–1981.

[14] Yang Liu, Yibo Li, and Xiaofei Ji, "Abnormal event detection in nature settings," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 7, no. 4, pp. 115–126, 2014.

[15] Dawei Du, Honggang Qi, Qingming Huang, Wei Zeng, and Changhua Zhang, "Abnormal event detection in crowded scenes based on structural multi-scale motion interrelated patterns," in *Multimedia and Expo (ICME), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1–6.

[16] Yang Hu, Yangmuzi Zhang, and Larry Davis, "Unsupervised abnormal crowd activity detection using semi-parametric scan statistic," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 767–774.

[17] Ying Zhang, Huchuan Lu, Lihe Zhang, and Xiang Ruan, "Combining motion and appearance cues for anomaly detection," *Pattern Recognition*, vol. 51, pp. 443–452, 2016.

[18] Mehrsan Javan Roshtkhari and Martin D Levine, "Online dominant and anomalous behavior detection in videos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2611–2618.

[19] Mohammad Sabokrou, Mohsen Fayyaz, Mahmood Fathy, and Reinhard Klette, "Deep-cascade: Cascading 3d deep neural networks for fast anomaly detection and localization in crowded scenes," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1992–2004, 2017.

[20] Tan Xiao, Chao Zhang, and Hongbin Zha, "Learning to detect anomalies in surveillance video," *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1477–1481, 2015.

[21] Oren Boiman and Michal Irani, "Detecting irregularities in images and in video," *International journal of computer vision*, vol. 74, no. 1, pp. 17–31, 2007.