AN EFFICIENT ANOMALY DETECTION APPROACH IN SURVEILLANCE VIDEO BASED ON ORIENTED GMM

Feiping Li,Wenming Yang, Qingmin Liao*

Graduate School at Shenzhen, Tsinghua University, China
Department of Electronic Engineering
Shenzhen Key Lab. of Information Sci& Tech/Shenzhen Engineering Lab. of IS& DRM

ABSTRACT

The detection and localization of abnormal activities are considered in this work. An efficient approach called oriented G-MM(OGMM) is proposed. The approach uses optical flow as low-level feature and quantizes the orientation of optical flow into 8 sections. In training stage, the approach will learn a GMM model at each orientation section and each position. In testing stage, the proposed approach estimates the probability of whether a position is abnormal using likelihood method. The proposed approach is a local method and can detect and locate anomaly. What's more, in the proposed approach, the same process is done to each position with little interaction between different positions. This makes the approach suit for parallel computing and can deal with large-scale tasks in Big Data times. The experiments verify that the proposed approach is efficient and effective.

Index Terms— optical flow, anomaly detection, oriented GMM, Big Data

1. INTRODUCTION

Video surveillance has been widely used in recent years. In future, when the Big Data times comes, the video surveillance may be in everywhere around us. However, so far, the videos are only surveyed after some accidents having happened. A lot of loss may have appeared already when surveillance videos contribute. Or in other case, it needs human to stare at the screens all the time, which may lead to tiredness and inattention. It's obvious that this way is inefficient and will even hard to work in Big Data times. Under this circumstance, there is a great need for an effective abnormal behaviour detection approach.

As a result, in recent years, many researchers have paid their attention to this area. Many strategies are proposed, which contain both tracking-based approaches and non-tracking approaches. The tracking-based approaches [1–4] are based on moving objects detection and tracking algorithms, and may have good effects in scenes with few people. But in scenes with a lot of moving objects, they will fail to track and suffer heavily decrease of performance.

The non-tracking approaches mainly use technologies in statistics [5–14], and will have better performance in scenes with crowded people. Among these researches, [6, 7, 9–16] use optical flow as the feature, among which some of them [9, 10, 12] use the histogram of optical flow as the feature descriptor. Some of works use methods in statistics learning. For example, [6, 11] use the mixtures of dynamic textures method, and [7] uses the space-time Markov Random Field(MRF) model to detect temporal and spatial anomalies. But their methods are time-consuming and have not satisfying performance. Some other works use the sparse reconstruction method [5, 15, 16],or the low-rank matrix approximation [17] and so on. They use the representation error to detect abnormal activities. These methods are also very time-consuming, because they need to represent every pixel or every patch of every frame. In terms of speed of processing, [18] introduced a very fast method to detect abnormality in Compressed Domain. But the method depends on video compression algorithm and can't design anomaly detection algorithm freely. So its performance is limited.

The proposed approach is non-tracking based. In our approach, the optical flow feature is used. The optical flow will be quantized into 8 sections based on its orientation. The approach will train a GMM model at each orientation section and each position. In testing stage, the likelihood strategy is used to produce an anomaly likelihood map for each frame. The map will be used to determine whether an anomaly is happened. In order to reduce false alarms, a local mean blur operation in spatial domain is utilized, coming together with a minimum filter operation in temporal domain. The two operations are all implemented on the likelihood map sequence. They can suppress small and short clutter spots.

Our proposed approach is an efficient statistics learning methods. It can detect anomaly in velocity, orientation and position. The proposed approach has several advantages. First, it is very fast and efficient, because the dense grids

This work was supported by NSFC under Grant No.61471216 and Special Foundation for the Development of Strategic Emerging Industries of Shenzhen under Grant No.ZDSYS20140509172959974.

*Corresponding author: yangelwm@163.com (W. Yang)

method is used to improve the speed of processing, without decrease in performance. And it suits for parallel computing and stream processing. Second, it needs only a little initial data for it can incrementally update parameters online.

2. THE PROPOSED APPROACH

Our proposed approach is inspired by the widely-used GMM method. It can be called oriented GMM method. The overview of our approach is illustrated in Fig.1(a). In the training stage, we first compute optical flow field of every frame. Considering of speed of processing, we use dense grids to represent the whole frame. Each grid node stand for a position in the scene. The positions(grid nodes) whose magnitude of optical flow are under certain threshold are considered as background, and will not be considered as abnormal. Then the optical flow of the foreground will be quantized into 8 sections, according to the orientation. The magnitude of optical flow of same position and same orientation section is considered to follow the mixture of Gaussian distributions.

2.1. Preprocess

Given continuous video frame sequence, after computing optical flow field of every frame, we will produce a dense grid. Optical flow of each grid node is computed from its surrounding patch, using the same method as [13]. All the pixels in a patch are assigned into 8 categories according to the orientation of optical flow as shown in Fig.1(b). The orientation of corresponding grid node is determined to be the most frequent orientation of pixels in the patch. And the magnitude of the grid node is determined to be the average of pixels in the patch. But the pixels in the background will be ignored. Obviously, optical flow of grid node is more robust than pixels.

2.2. Oriented GMM

It's reasonable to assume that the velocity of normal walk of humans at given position and orientation in a 3-D scene follows the mixture of Gaussian distributions. When the 3-D scene is projected into the 2-D image plane, the velocity of humans is also projected as optical flow. The relationship between the magnitude of the velocity and the magnitude of optical flow can be expressed as follows:

$$V_{OF} = \frac{V}{d}\sin\theta \tag{1}$$

Where V and V_{OF} denote the magnitude of the velocity and optical flow respectively. d denotes the distance between the position and the camera. θ denotes the angle between the real velocity and the optical axis of the camera.

For fixed surveillance cameras, the d of certain position is constant. The θ is also constant for given orientation of optical flow. This can be explained as follows: For surveillance video, human can only walk on ground. And usually, the optical axis of the surveillance cameras do not parallel the ground, because most of the cameras are installed at high places in order to reduce occlusion. So the orientation of the real velocity and optical flow are one-to-one correspondence.

According to our analysis above, at given position and orientation, the magnitude of optical flow V_{OF} is similar to V and also follows the mixture of Gaussian distributions. So we propose the oriented GMM approach. After the quantization of the orientation of optical flow in preprocess section, we will train a GMM model at each orientation section and each position.

2.3. Training stage

According to the analysis above, a mixture of Gaussian distributions is needed at each position(each grid node) and each orientation. Assume the number of Gaussian components is K. For each grid node that is not background, it is first mapped to corresponding mixture of Gaussian distributions depending on its position and optical flow orientation. It is then assigned to corresponding Gaussian component depending on its probabilities belonging to each component P_k. For grid node that is background, the corresponding GMM will be freezed.

$$P_k = e^{-\frac{||V_{OF}-M_k||^2}{2Var_k}} \tag{2}$$

Where M_k and Var_k are the mean and variance of kth Gaussian component respectively. The V_{OF} is the the magnitude of optical flow. The node will be assigned to the Gaussian component with maximum probability. Then the parameters of the Gaussian component will be updated using equation (3).

$$\begin{cases} M_k^{N+1} = \frac{N}{N+1}M_k^N + \frac{1}{N+1}V_{OF} \\ Var_k^{N+1} = \frac{N}{N+1}Var_k^N + \frac{1}{N+1}(V_{OF} - M_k^{N+1})^2 \\ N = N+1 \end{cases} \tag{3}$$

Where N denotes the total number of nodes belonging to the Gaussian component until current time.

In this way, we can update the parameters of all mixtures of Gaussian distributions incrementally, only with some reasonable initial values. This feature is important, because we don't need store large volume of video data in advance.

2.4. Testing stage

In testing stage, a similar process with the training stage is implemented. Firstly, the grid nodes whose magnitude of optical flow under certain threshold are considered as background. These nodes are impossible to be abnormal. Then, for each node in foreground, it is mapped to corresponding mixture of Gaussian distributions depending on its position and optical flow orientation. After that, it is assigned to corresponding Gaussian component depending on its probabilities

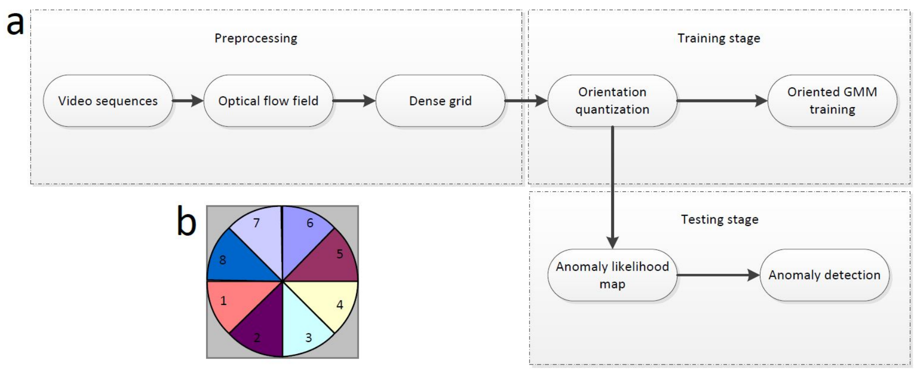

Fig. 1. (a). The overview of our approach. (b) Quantization of the orientation of optical flow

belonging to each component.

$$k_m = \arg\max_k P_k \tag{4}$$

Where P_k is calculated using equation (2). Then the likelihood Lp will become the indicator of whether a position is abnormal or not.

$$p = -\log P_{k_m} \tag{5}$$

In order to suppress small and short clutter spots, a local mean blur operation in spatial domain is utilized, coming together with a minimum filter operation in temporal domain. The operation is described in equal (6).

$$\begin{aligned} \overline{Lp_{t,j}} &= \tfrac{1}{N_j} \sum_{i \in N(j)} Lp_{t,i} \\ Lp^*_{t,j} &= \min(\overline{Lp_{t,j}}, \overline{Lp_{t-1,j}}) \end{aligned} \tag{6}$$

Where $Lp_{t,i}$ denotes the likelihood of position i at time t.

It's alternative whether the parameters of Gaussian distributions to be updated or not in the testing stage. If we choose to update the parameters, the equation (3) is used. It need to be note that we can update parameters both in training stage and testing stage. But they are different. In training stage, there are only normal data, so we can use all of them. But in testing stage, we can only use the normal data we have detected.

From the preceding statement about the proposed approach, we can find that the approach can detect anomaly in velocity. As about anomaly in orientation and position, we need only initialize the parameters of all mixture of Gaussian distributions with a very small value. The parameters of mixture of Gaussian distributions corresponding to abnormal orientation and position will not be updated in training stage and will keep small. So activities in orientation and position will have very big likelihood of being abnormal according to equation (2),(4) and (5). This makes the anomaly distinguishable.

And from the description of Section 2, it can be found that the proposed approach is a local method. In the approach, same process is done to each position. And there are only neighbouring operations between different positions. This makes the approach fit to parallel computing and can deal with large-scale tasks. For examples, with parallel computing, we can deal with lots of channels of video with only one machine. And meanwhile, the proposed method is an online method. There are only neighbouring operations between different frames. So that, our approach need not store many historical frames and suits for stream processing. This can save the cost of big volume memory and storage.

3. EXPERIMENT

The experiments are implemented on a widely used dataset, the UCSD anomaly detection dataset [6]. It contains two subsets, ped1 and ped2 dataset. The ped1 subset contains 34 training sequences and 36 testing sequences. Each sequence consists of 200 frames. The ped2 subset contains 16 training sequences and 12 testing sequences. There are 120,150 or 180 frames in each sequence.

The recognition rate (RD) of several comparison approaches and our approach on ped1 subset with pixel-level ground-truth are listed in Table 1. Among these approaches, the SF represents the Social Force approach proposed in [8]. The Mixture of Probabilistic Principal Component Analyzers approach proposed in [7] is written as MPPCA. SF+MPPCA is the short name of the Social Force with MPPCA approach used as comparison in [6]. The approach proposed in [19] has not appropriate abbreviation. It is written as its first author's name, Adam *et al*. MDT stand for the Mixture of Dynamic Texture approach proposed in [6]. The newly proposed approach in [5], which uses the sparse coding method, is also written as its first author's name, Li *et al*. The experiments on ped2 subset are also done. But the results are not listed out, because there are no comparing results of recognition rate (RD)in other literatures. [11] has a superior results, mainly

Fig. 2. Some results of the proposed approach in locating anomaly.

Table 1. RD(%) pixel-level recognition rate on ped1

Algorithem	SF	MPPCA	SF+MPPCA	Adem *et al*	MDT	Li *et al*	Ours
RD	21	18	28	24	45	63	57

because it uses a CRF filter after the detection of anomaly. Its main detecting method is still MDT, as same as [6].

Some of the results of detecting and localizing the abnormal activities are shown in Fig.2. And the ROC curves of the recognition rate of our approach and comparison approaches are shown in Fig.3. In the experiments, the number of Gaussian components K is set to 1. Actually, the recognition rate changes tiny when the value of K changes from 1 to 6. And with different computing method of optical flow, the K need to be different value to get the best performance.

From Fig.3, it can be found that the RD of approach

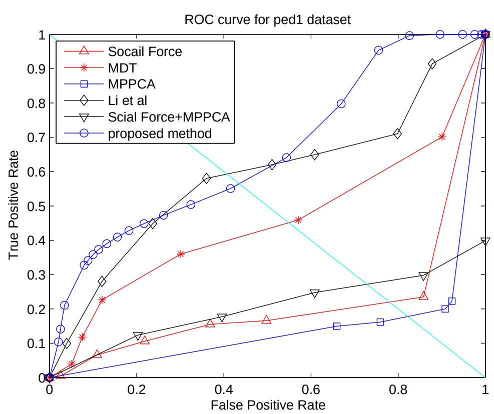

Fig. 3. ROC curves of our approach and comparison approaches on Ped1 Dataset with pixel-level ground truth.

used by Li *et al* is a little better than us, but its area under ROC (AUC) is worse. This means the robustness of our approach is better. And what's more, the MDT approach and the approach of Li *et al* is complex and time-consuming. The former uses over 25s per frame in testing stage, and the latter needs 3.4s/frame for ped1 dataset and 4.8s/frame for ped2 dataset.

Table 2. the time cost of our approach(s/frame)

Stage	Training	Testing
Ped1	0.03s	0.04s
Ped2	0.06s	0.06s

The time our approach used is listed in Table 2. Our experiments are implemented with MATLAB on a similar machine (2.67GHz CPU and 2GB RAM) with Li *et al* [5](3GHz CPU and 2GB RAM) and [6](the MDT approach,3GHz CPU and 2GB RAM). It can be found that our approach is much faster.

4. CONCLUSION

In this work, a new abnormal activity detection approach called oriented GMM(OGMM) is proposed. The approach uses optical flow as low-level feature and quantizes the orientation of optical flow into 8 sections. And at each orientation section and each position in the given scene, we train a GMM model to learn the normal patterns. Then in testing stage, we use likelihood method to determine whether a position is abnormal. The OGMM is a local approach which can easily localize the position of abnormal with the characteristic of updating incrementally. The approach suits for parallel computing and stream processing. It is a kind of approach fit to use in Big Data times.

The proposed approach is tested on a newly published dataset, the UCSD abnormal detection dataset. The results have verified that our approach is fast and effective. Its performance exceeds most of the recent approaches.

5. REFERENCES

[1] XiaoBin Zhu, Xin Jin, XiaoYu Zhang, ChangSheng Li, FuGang He, and Lei Wang, "Context-aware local abnormality detection in crowded scene," *Science China Information Sciences*, vol. 58, no. 5, pp. 1–11, 2015.

[2] Jing Cui, Weibin Liu, and Weiwei Xing, "Crowd behaviors analysis and abnormal detection based on surveillance data," *Journal of Visual Languages & Computing*, vol. 25, no. 6, pp. 628–636, 2014.

[3] Xinyi Chong, Weibin Liu, Pengfei Huang, and Norman I Badler, "Hierarchical crowd analysis and anomaly detection," *Journal of Visual Languages & Computing*, vol. 25, no. 4, pp. 376–393, 2014.

[4] Yuan Yuan, Jianwu Fang, and Qi Wang, "Online anomaly detection in crowd scenes via structure analysis," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 562–575, 2015.

[5] Nannan Li, Xinyu Wu, Dan Xu, Huiwen Guo, and Wei Feng, "Spatio-temporal context analysis within video volumes for anomalous-event detection and localization," *Neurocomputing*, vol. 155, pp. 309–319, 2015.

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

[7] Jaechul Kim and Kristen Grauman, "Observe locally, infer globally: a space-time mrf for detecting abnormal activities with incremental updates," in *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009*. IEEE, 2009, pp. 2921–2928.

[8] Ramin Mehran, Akira Oyama, and Mubarak Shah, "Abnormal crowd behavior detection using social force model," in *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009*. IEEE, 2009, pp. 935–942.

[9] Tian Wang and Hichem Snoussi, "Detection of abnormal events via optical flow feature analysis," *Sensors*, vol. 15, no. 4, pp. 7156–7171, 2015.

[10] 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.

[11] 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.

[12] Lijun Wang and Ming Dong, "Detection of abnormal human behavior using a matrix approximation-based approach," in *Machine Learning and Applications (ICMLA), 2014 13th International Conference on*. IEEE, 2014, pp. 324–329.

[13] Dong-Gyu Lee, Heung-Il Suk, and Seong-Whan Lee, "Modeling crowd motions for abnormal activity detection," in *11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2014*. IEEE, 2014, pp. 325–330.

[14] Myo Thida, How-Lung Eng, and Paolo Remagnino, "Laplacian eigenmap with temporal constraints for local abnormality detection in crowded scenes," *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 2147–2156, 2013.

[15] Xun Tang, Shengping Zhang, and Hongxun Yao, "Sparse coding based motion attention for abnormal event detection," in *20th IEEE International Conference on Image Processing (ICIP), 2013*. IEEE, 2013, pp. 3602–3606.

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

[17] Lijun Wang and Ming Dong, "Real-time detection of abnormal crowd behavior using a matrix approximation-based approach," in *19th IEEE International Conference on Image Processing (ICIP), 2012*. IEEE, 2012, pp. 2701–2704.

[18] Huang Li, Yihao Zhang, Ming Yang, Yangyang Men, and Hongyang Chao, "A rapid abnormal event detection method for surveillance video based on a novel feature in compressed domain of hevc," in *IEEE International Conference on Multimedia and Expo (ICME), 2014*. IEEE, 2014, pp. 1–6.

[19] Amit Adam, Ehud Rivlin, Ilan Shimshoni, and David Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 555–560, 2008.