ANOMALY DETECTION IN SURVEILLANCE VIDEO USING MOTION DIRECTION STATISTICS

Chang Liu, Guijin Wang, Wenxin Ning, Xinggang Lin

Tsinghua University
Department of Electronic Engineering
Beijing 100084, China

Liang Li, Zhou Liu

Sony China Research Lab
Beijing 100190, China

ABSTRACT

A novel approach for detecting anomaly in visual surveillance system is proposed in this paper. It is composed of three parts:(a) a dense motion field and motion statistics method, (b) one-class SVM for one-class classification, (c) motion directional PCA for feature dimensionality reduction. Experiments demonstrate the effectiveness of proposed algorithm in detecting abnormal events in surveillance video, while keeping a low false alarm rate. Moreover, it works well in complicated situation where the common tracking or detection module won't work.

Index Terms— Visual surveillance, Anomaly detection, Motion vector, One-class SVM, PCA

1. INTRODUCTION

Security problems are attracting increasingly more attention nowadays. Millions of surveillance cameras are placed to discover abnormal events, which creates the premise that an alarm can be released automatically when abnormal events occur. Therefore, developing an anomaly detection algorithm is of significant importance.

Numerous efforts have been made to detect anomaly in surveillance video. Related works fall into two categories briefly. The first category contains methods based on object detecting or tracking, and the second category contains methods based on classification using low-level features.

Based on detection or tracking results, methods in the first category extract trajectory features [1] [2] or make semantic analysis [3] [4] [5] to detect anomaly. For instance, [1] addresses the problems of track matching and dynamic event detection in a sequence of frames. In paper [2], trajectory clustering is used to distinguish between normal and abnormal events. Cui et.al. [4] use the probability of an observation with each event state to estimate prior and posterior state distribution, as well as sequential Monte Carlo framework extended by Markov Random Field for tracking interactive events. However, surveillance scenarios usually have complex occlusion, complex illumination condition and low resolution. Therefore, precise detection or tracking of individual human is very difficult.

In the second category, low-level features such as shape, position or motion are utilized. Wang et.al. [6] treat the anomaly detection as a human shape recognition problem. But it depends on accurate human edge detection. In paper [7], a novel statistical framework is presented for modeling local spatio-temporal motion pattern. It deals with moving in reverse or irregular patterns in an extremely crowded scene. In Ihaddadene's paper [8], the result is measured by scalar product of the normalized values of several factors calculated by motion vector. They successfully detect collapsing events at an airport escalator exits. In [9], local motions in four directions are used as low level features, they discovering behaviours by using their dynamics. [10] use motion pattern in traffic scene to detect the anomaly. The above works [7] [8] [9] [10] use motion as feature to detect anomalies, but they only detect a special pattern of anomaly rather than the opposite of normal events. In [11], binary foreground image is employed as features, and one-class SVM [12] is used to train an anomaly detector. A merit of one-class SVM is that only normal samples are required. It is especially suitable for ATM or similar scene. However, it is sensitive to position and does not perform very well in busy environments such as a hall or plaza.

In this paper, we propose an anomaly detection system based on low-level features. Dense motion field and statistics are computed in each frame. Then motion directional PCA technique is presented to extract useful principle features in time-span. Finally one-class SVM discriminates the anomaly from normal events. Experimental results show that our scheme can effectively detect the anomaly events with a low false alarm rate.

2. FRAMEWORK

In order to train an anomaly detector, we need to prepare samples and corresponding features. The abnormal event happens very rarely and so is hard to obtain. In addition, unlike the normal walking videos in feature space, every abnormal video is different from the others. Fortunately, only the normal case samples are needed for training the detector by one-class SVM.

Our proposed algorithm consists of a detector training part and a detection part. We first compute motion vector of blocks in every frame. Then a statistics method is used to obtain a motion directional statistics features. After we obtain the features for a set of frames, a dominant direction of these features can be derived. We only use videos containing normal scenes for training. The block diagram in Figure 1 shows the training process. A motion directional PCA is operated on the features, followed by training a detector using the one-class SVM [12] which describes the domain of normal samples. In the detection procedure(Figure 2), features are obtained in the same manner, and then classified by one-class SVM to get detection results. The event will be determined abnormal if the features do not fit the domain description.

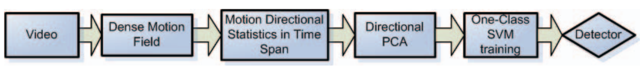


Fig. 1. Framework of training algorithm.

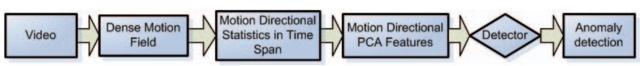


Fig. 2. Framework of detection process.

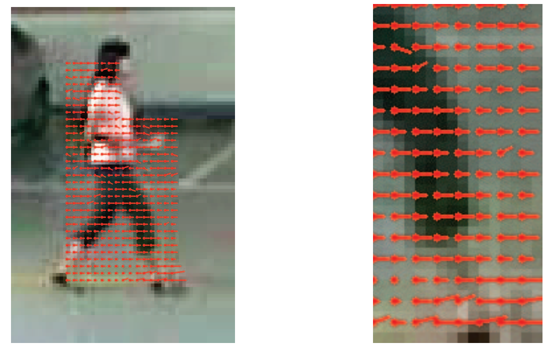
3. MOTION VECTOR STATISTICS FEATURE

This section describes an effective feature extracting method constructed by block motion vector. Many existing approaches extract features based on the magnitude and direction of the motion vectors in a frame [7] [8] [9] [10], which can reflect the characteristics of object movements, especially with fixed camera.

3.1. Dense motion field

One weakness of using motion vector statistics for finding movement is the lack of motion vector.

In conventional technologies like video compression algorithm, an image frame is divided into many k $\times$ k blocks and then the motion vector of each block is calculated, which provides insufficient amount of useful motion vectors from one frame, especially in surveillance video. For example, in a 640 $\times$ 480 sized video in underground parking lot, a typical pedestrian in the video only corresponds to about 8 motion vectors. Insufficient motion vectors will lead to incorrect directional statistics. Grid effect is also significant and bad for directional statistics. Therefore, a dense motion field is needed.



(a) Part of dense motion field (b) Part of dense motion field

Fig. 3. Dense motion field.

To get more motion vectors, overlapping blocks are adopted. In practice, we use 2-pixels shift both horizontally and vertically. By doing this, we get 64 times of motion vectors as before, as shown in Figure 3. This improvement makes the statistics more reliable.

3.2. Motion vector directional statistics

The motion vectors cannot be used as features directly. Instead, we use motion vector statistics as features. A statistics is first operated on the direction of all the blocks. The directional features are then derived from the set of motion vectors $V_{i,j}$. Let

$$|V_{i,j}| = \sqrt{(V_{i,j}^x)^2 + (V_{i,j}^y)^2} \tag{1}$$

$$\Phi_{i,j} = \arctan(V_{i,j}^y / V_{i,j}^x) \tag{2}$$

respectively denote the amplitude and direction of the motion vector $V_{i,j}$.

We are more concerned about direction, because the major feature of object in surveillance video is motion direction. Under normal circumstances, a pedestrian usually moves towards a certain destination with little deviation or rotation movement. On the contrary, in abnormal condition, people do not have a constant direction and move in disorder. Therefore motion vector direction can be employed to distinguish between normal and abnormal events.

Moreover, conventional directional statistics divided the angles into individual sections. However, many motion vectors occur at the boundary between two sections. Therefore, we use overlapped sections instead.

4. MOTION DIRECTIONAL PCA

In this section, an effective feature dimensionality reduction method by motion directional PCA is described.

4.1. Feature Series in time-span

With the motion directional features in a single frame, it is still hard to distinguish between abnormal events and normal walking, especially when several people are walking. Therefore, we need to combine a sequence of motion directional features. In practice, we use motion field statistics in consecutive T frames to form feature vectors. We first extract a main direction in each frame, then find the dominant direction for the entire sequence and set this as the first element in the feature vector for each frame. In this way, we can make the feature vectors direction invariant.

4.2. Motion Directional PCA

After obtaining the motion vector statistics of several normal videos, we will use these features to train a detector. The dimension of feature is set to the number of frames in a sample times the direction number. However, such high-dimensional features will degrade the efficiency of training one-class SVM [12] due to the large amount of redundant dimensions.

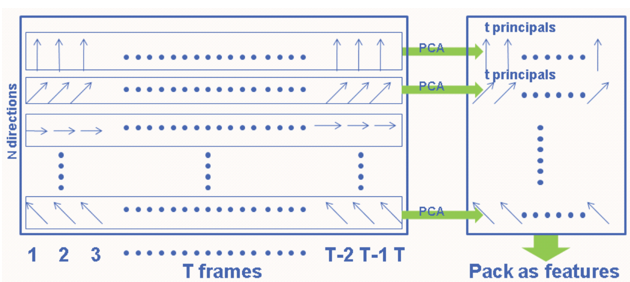


Fig. 4. Motion directional PCA.

A modified PCA (principle component analysis) [13] is adopted on the features to make them more compact as well as informative. Instead of using traditional PCA directly on high-dimensional features, we apply PCA to each separate direction of the features, keeping only the top 95% energy of each direction in the result. As shown in Figure 4, the PCAs on different direction is independent. T is the frame number of the detection time window. Every PCA pack the motion vector statistics features in the same direction into a smaller feature vector. We call this motion directional PCA.

After integrating the motion directional PCA into the framework, we use single-walking and multi-walking as normal samples, and succeed in distinguishing anomaly from the normal ones.

5. EXPERIMENTS

5.1. Datasets

We used two datasets to test our algorithm. In the first dataset, PETS2004 dataset, the camera looks downward using a wide angle camera lens. The resolution is 384×288, with 25 frames per second. It includes actions such as walking, browsing, resting, meeting, walking together and splitting up and two people fighting.

Fig. 5. Example frames of Underground parking dataset.

As shown in Figure 5, the second dataset we used is from an underground parking lot surveillance video. The resolution is 640×480, with 30 frames per second. The illumination conditions are much worse than the PETS2004 dataset. In a two-and-a-half-hours duration, events including single-walking, group-walking, talking, meeting and fighting were captured.

5.2. Results of anomaly detection

We manually labled the abnormal period in the video as ground truth. The beginning of the period is set as the first physical contact of the opposite sides. The end of the period is set as one of the opposite sides leave the visible region. A time domain filtering is used on the results of detection module.

Under this criterion, we evaluated our algorithm in the above two datasets. A frame number of T=50 is used. In the PETS2004 dataset(Figure 6), there are 24 normal videos including walking, browsing, leaving bags, meeting and walking together and splitting up and 4 abnormal videos of two men fighting. We use 7 normal videos as training samples. We succeed in detecting 3 fight video with no false alarm in the other 17 normal videos. The only incorrectly rejected abnormal video is $Fight\_Chase.mpg$ in which the fighting only last for a very short duration and the limb movements are not intensive. The results are summarized in Table 1.

In the Underground Parking dataset(Figure 7), there are 17 fighting events in the two-and-a-half-hour duration. We use 25 minutes of the video without fighting as normal samples. After training the anomaly detector, we test it on the whole video. It detects 15 fight events. And also, there are about 3-minute false alarms in the 140 minutes normal duration. Most of the false alarm is caused by the intense illumination change when vehicle headlamps are turned on, which seriously affects the accuracy of motion vector. The results are summarized in Table 2.

Fig. 6. Example frames of precise anomaly detection in PETS2004 dataset.

Fig. 7. Example frames of precise anomaly detection in Underground Parking dataset.

6. CONCLUSION

In this paper, we propose an anomaly detection method in surveillance video. Our algorithm employs a feature of motion vector statistics. The proposed feature is easy to calculate and not rely on any detection or tracking module, making the system easier to handle event detection with occlusion. A motion directional PCA is applied for feature down dimension, which is important for the efficiency of subsequent training. One-class SVM is employed to overcome the problem of lacking abnormal data. Experimental results show that most abnormal events are detected by the proposed algorithm with a very low false alarm rate.

7. REFERENCES

[1] T. Ding, M. Sznaier, and O. I. Camps, "Fast track matching and event detection," in *CVPR*, 2008, pp. 1–8.

[2] F. Jiang, Y. Wu, and A. K. Katsaggelos, "Abnormal event detection based on trajectory clustering by 2-depth greedy search," in *ASSP*, 2008, pp. 2129–2132.

[3] F. Cupillard, A. Avanzi, F. Bremond, and M. Thonnat, "Video understanding for metro surveillance," in *Proc. IEEE International Conference on Networking, Sensing and Control*, 2004, vol. 1, pp. 186–191.

Table 1. PETS2004 Dataset results

Results	Detect as Abnormal	Detect as Normal	Precise Rate
Normal	0s	387s	100%
Abnormal	17s	3s	85%

Table 2. Underground Parking Dataset results

Results	Detect as Abnormal	Detect as Normal	Precise Rate
Normal	180s	8268s	98%
Abnormal	364s	46s	89%

[4] P. Cui, L. Sun, Z. Liu, and S. Yang, "A sequential monte carlo approach to anomaly detection in tracking visual events," in *CVPR*, 2007, pp. 1–8.

[5] W. Lao, J. Han, and PHN de With, "Automatic surveillance analyzer using trajectory and body-based modeling," in *Proc. Digest of Technical Papers International Conference on Consumer Electronics*, 2009, pp. 1–2.

[6] Y. Wang, K. Huang, and T. Tan, "Human activity recognition based on r transform," in *CVPR*, 2007, pp. 1–8.

[7] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," in *CVPR*, 2009, pp. 1446–1453.

[8] N. Ihaddadene and C. Djeraba, "Real-time crowd motion analysis," in *ICPR*, 2008, pp. 1–4.

[9] T. Hospedales, S. Gong, and T. Xiang, "A markov clustering topic model for mining behaviour in video," in *ICCV*, 2009.

[10] I. Saleemi, L. Hartung, and M. Shah, "Scene understanding by statistical modeling of motion patterns," in *CVPR*, 2010.

[11] K. Sudo, T. Osawa, K. Wakabayashi, H. Koike, and K. Arakawa, "Estimating anomality of the video sequences for surveillance using 1-class svm," *IEICE - Trans. Inf. Syst.*, vol. E91-D, no. 7, pp. 1929–1936, 2008.

[12] DMJ Tax and RPW Duin, "Support vector domain description," *Pattern Recognition Letters*, vol. 20, pp. 1191–1199, 1999.

[13] A. K. Jain, *Fundamentals of digital image processing*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1989.