PAPER

Drastic Anomaly Detection in Video Using Motion Direction Statistics

Chang LIU[†], ***Student Member***, **Guijin WANG**[†a)], ***Member***, **Wenxin NING**[†], ***and*** **Xinggang LIN**[†], ***Nonmembers***

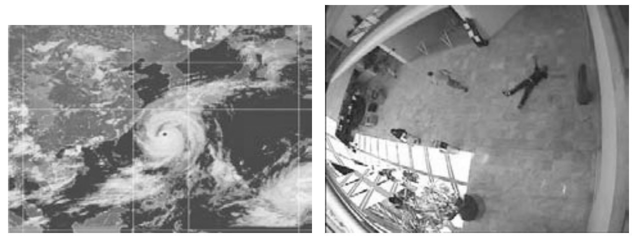
SUMMARY A novel approach for detecting anomaly in visual surveillance system is proposed in this paper. It is composed of three parts: (a) a dense motion field and motion statistics method, (b) motion directional PCA for feature dimensionality reduction, (c) an improved one-class SVM for one-class classification. Experiments demonstrate the effectiveness of the proposed algorithm in detecting abnormal events in surveillance video, while keeping a low false alarm rate. Our scheme works well in complicated situations that common tracking or detection modules cannot handle.
key words: *visual surveillance, anomaly detection, motion vector, one-class SVM, PCA*

1. Introduction

Detecting anomalies plays a crucial role in the understanding and interpretation of visual phenomenon. It has many applications, such as emergency detection of indoor or outdoor visual surveillance and satellite cloud image analysis. Figure 1 illustrates two applications of anomaly detection. In Fig. 1 (a), by analyzing the sequence of the cloud image, some hazardous weather (typhoons, severe convective weather or tropical storm) can be detected automatically. In Fig. 1 (b), in the surveillance video, an ideal analysis can get a lot of useful information, including the prohibition area alarm, exception event alarm or customary path analysis etc.

Numerous efforts have been made to detect anomaly in visual phenomenon. Related works fall into two categories roughly. The first category analyzes the abnormal behavior based on object detection and tracking; the second category observes the anomaly based on statistical low-level features.

Methods in the first category extract trajectory features [1], [2] or do semantic analysis [3]–[7] to detect anomaly using detection or tracking results as foundation. Ding et al. [1] addressed the problems of track matching and dynamic event detection in a sequence of frames. Trajectory clustering [2] was used to discriminate between normal and abnormal events. On the other side, with the help of co-occurrence matric, spatial histogram of the detected objects was also employed as features to detect unusual activities unsupervised [4]. Spatial-temporal volumetric features [5] were used to efficiently scan video sequences in space and time. Cui et al. [6] used the probability of an observation with each event state to estimate prior and posterior state distribution, as well as sequential Monte Carlo framework



(a) Satellite cloud image (b) Visual surveillance

Fig. 1 Typical applications of anomaly detection.

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. Some researchers treated the anomaly detection as a human shape recognition problem [8]. But it depended on accurate human edge detection, which is very difficult to achieve in surveillance video. A novel statistical framework was presented for modeling local spatio-temporal motion pattern [9]. It dealt with moving in reverse or irregular patterns in an extremely crowded scene. Some researchers also used motion direction as base feature and used hierarchical Bayesian models to do activity estimation [10]. Mehran [11] exploited abnormal crowd behavior to search and analyze the different movements in the surrounding. In Ihaddadene's research [12], the result was measured by scalar product of the normalized values of several factors calculated by motion vector. They successfully detected collapsing events at an airport escalator exit. Among these features, motion is one of the most widely used features.

The above works [9]–[12] used motion as feature to detect anomalies, but they only detected a special pattern of anomaly rather than the opposite of normal events. Some pioneering works were in the direction of these. For example, some researchers [13] dealt with the anomaly as all the events other than normal ones. Binary foreground image is employed as features, and one-class SVM [14] was used to train an anomaly detector. One advantage of one-class SVM is that only normal samples are required. It is especially suitable for ATM or similar scenes. It can handle fight events as well as peep events. However, it is sensitive to position and does not perform very well in busy environments such as a hall or plaza. Lin et al. [15] used integral motion

Manuscript received December 2, 2010.
Manuscript revised March 28, 2011.
[†]The authors are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China.
a) E-mail: wangguijin@tsinghua.edu.cn
DOI: 10.1587/transinf.E94.D.1700

of separate objects with Lie algebraic representation to describe the motion pattern. The outlier can be detected by the model afterward.

Anomaly include several categories that are extremely different from each other. So using a common method to detect all anomaly is impossible. In the literature, there is no universally accepted anomaly classification method. But for a solution by computer, a strict classification is essential. We classify the anomaly into two types: drastic anomaly and calm anomaly. Drastic anomaly is the anomaly with a large number of unusual movement. The calm anomaly means anomaly with no special movement. In this paper, we focus on drastic anomaly. We employ motion direction statistics feature to detect anomaly in surveillance video. One-class SVM is utilized to train a detector. A motion directional PCA method is proposed to extract useful principles from gross features. In other words, using motion directional PCA and one-class SVM, we use low level features to achieve anomaly detection which is a high level description of surveillance video.

The rest of the paper is organized as follows: Section 2 describes the framework of our method. Section 3 gives the foundation of motion field statistics features. The proposed motion directional PCA algorithm is presented in Sect. 4. The method of features summing up in time span and the domain description is in Sect. 5. Some implementation details and discussion are in Sect. 6. Experiment results are presented in Sect. 7. Section 8 gives the conclusion.

2. Drastic Anomaly Detection 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 (Fig. 2) and a detection part (Fig. 3). We first compute motion vector of blocks in every frame. Then we use a statistics method to obtain a motion direction 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. A motion directional PCA is operated on the features, followed by training a detector using the one-class SVM which describes the domain of normal samples. In the detection procedure (Fig. 3), 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.

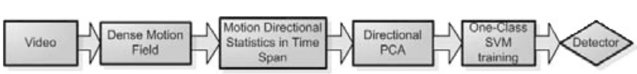


Fig. 2 Framework of training algorithm.

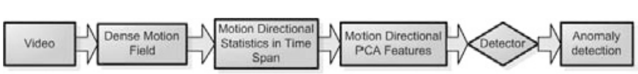


Fig. 3 Framework of detection process.

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 [9], [12], which can reflect the characteristics of object movements, especially with fixed camera.

3.1 Dense Motion Field

In conventional technologies like video compression algorithms, an image frame is divided into many k × 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 × 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 direction statistics. Grid effect also significantly degrade direction statistics. Therefore, a dense motion field is needed.

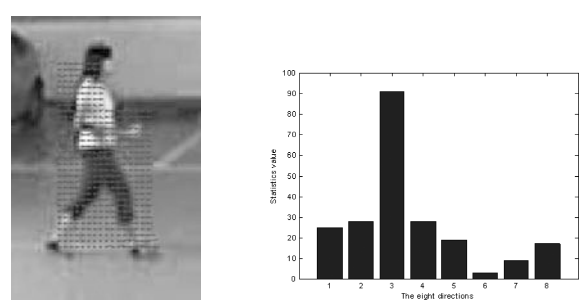
To get more motion vectors, overlapping blocks are adopted. In practice, we use 2-pixel shift both horizontally and vertically. By doing this, we get 64 times more motion vectors than before, as shown in Fig. 4. This improvement makes the statistics more reliable.

3.2 Motion Vector Direction Statistics

We use motion vector statistics as features. 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

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

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



(a) The dense motion field for a pedestrian
(b) Motion statistics in single frame

Fig. 4 Dense motion field.

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

We are more concerned about direction, because the major feature of objects in surveillance video is motion direction. For instance, a pedestrian usually moves towards a certain destination with little deviation or rotational movement. On the contrary, in abnormal condition, the motion field tends not to have a constant direction and move in disorder. Therefore motion vector direction can be employed to distinguish between normal and abnormal events. We divide angle of circumference into eight equal angles to make the statistics.

4. Motion Directional PCA

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. If we connect the features directly, the connotative organization in the high dimensional features is dissipated.

So an effective feature dimensionality reduction method by motion directional PCA is described.

4.1 Feature Series in Time-Span

In practice, we use motion field statistics in every 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. We set the rotational shift m as follows

$$m = \underset{0 \le i \le n-1}{\arg\max} \left(\sum_{t=0}^{T-1} D_{t,i} \right) \tag{3}$$

T is the frame number to form the feature vectors, and n is the direction number. $D_{t,i}$ is the statistics in the tth frame and the ith direction. We use this m to rotate the features, making the first feature column as the column with the greatest strength. (As shown in Fig. 5) In this example, every feature's sequence number reduces the number of two. In order to illustrate this process more clearly, the feature rotation of every single frame is shown in Fig. 6.

$$revised: D_{t,i} \ = D_{t,i-m} \tag{4}$$

The second subscript is defined as cycled by n, that means $D_{t,i}$ is the same as $D_{t,i+n}$. In this way, we can make the feature vectors direction invariant.

4.2 Motion Directional PCA

After obtaining the motion vector statistics, 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 feature will

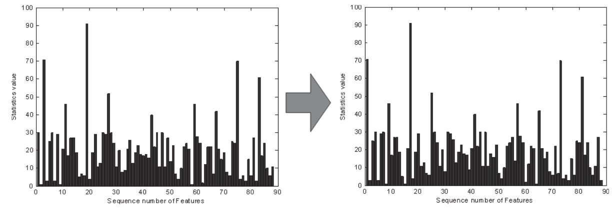


Fig. 5 Main direction arrangement.

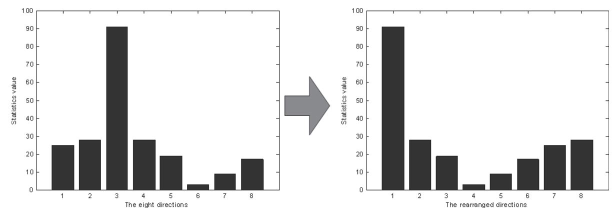


Fig. 6 Main direction arrangement in single frame.

degrade the efficiency of training one-class SVM, due to the large amount of redundant dimensions. We use PCA (principle component analysis) [16] on features before training.

Let $D = \{D_n \in R^d | n = 1, \ldots, N\}$ be an ensemble of feature vectors obtained from the video, where d is the original feature dimension. Let

$$E(D) = \frac{1}{N} \sum_{n=1}^{N} D_n \tag{5}$$

be the mathematical expectation of feature vectors. And the global covariance matrix M is calculated after subtracting the mathematical expectation.

$$M(D) = \frac{1}{N} \sum_{n=1}^{N} (D_n - E(D))(D_n - E(D))^T \tag{6}$$

$M(D)$ is a $d \times d$ matrix. It is proved that the matrix $M(D)$ is positive definite and has only real non-negative eigenvalues, then we can compute d positive eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_d$ with the corresponding eigenvectors $e_1, e_2, \ldots, e_d$ using Singular Value Decomposition. The first few eigenvectors contain higher energy of the training samples. Therefore we ignore some of the small eigenvalues to reduce the computation pressure on training detector. In order to determine how many eigenvectors to be ignored, we use a threshold value T_{energy}

$$P_k = \sum_{i=1}^{k} \lambda_i / \sum_{i=1}^{d} \lambda_i > T_{energy} \tag{7}$$

P_k is the energy proportion of the first k largest eigenvalues with respect to all eigenvalues. In our experiments, T_{energy} is chosen to be 95% to obtain best results.

Because the features are formed by the number of frames and the direction, we can take the feature vector more likely to a feature matrix (As shown in the left of Fig. 7). In practical, the motion has interdependency in time line, so we excavate the information from this.

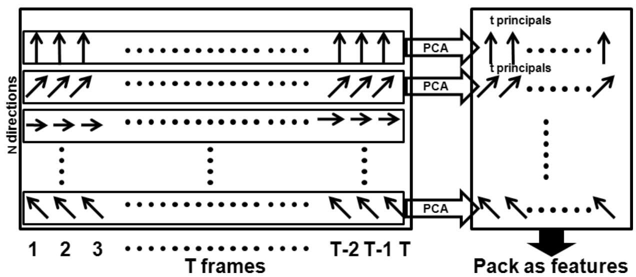


Fig. 7 Motion directional PCA.

A modified PCA is adopted on the features to make them more informative and compact. 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 Fig. 7, the PCAs on different directions are independent. Every PCA pack the motion vector statistics features in the same direction into a smaller feature vector. We call this motion directional PCA.

$$T_i = \{E_j D_{j,i}\}|_j = \{[e_{j,1}, e_{j,2}, \dots e_{j,k}]^T D_{j,i}\}|_j \tag{8}$$

where $j = 0, 1, 2, \dots, 7$.

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. Improved One-Class SVM

5.1 One-Class Classification

Generally speaking, with the absence of one side samples, the problem of one-class classification is harder than the problem of two-class classification. Some related efforts have been made before. Hanson et al. solved this problem of fitting a model to the data in [17]. Laaksonen et al. used a Self-Organizing Maps to do one-class classification in 1999 [18]. And Tax referred to a method called one-class SVM, inspired by the support vector machine [14]. They use a hypersphere instead of the hyperplane to define those one class samples. Among the above mentioned one-class classification methods, one-class SVM is resistant to the overfitting problem and model nonlinear relations in an efficient and stable way. When dealing with high dimensional feature and numerous samples, it performs exceptionally well.

5.2 One-Class SVM

The One-Class SVM [14] has been successfully applied to various learning problems [13]. It consists of learning the minimum volume contour that encloses most of the data in a dataset. Its original application is outlier detection, to detect data that differ from most of the data within a dataset. Let $X = \{x_i\}_{i=1}^n, x_i \in R^d$. Here, each x_i is the feature vector of a sample. The aim of One-Class SVM is to use the training data so as to learn a function $f_X : R^d - > R$, such that most of the data in X belong to the set $R_X = \{x | x \in R^d, f_X(x) > 0\}$ while the volume of R_X is minimal. It attempts to find a hypersphere in the feature space that separates the data from the outskirt space. This can be formulated into an optimization problem. We want the hypersphere to be as small as possible while at the same time, including most of the training data. We only consider the positive points and get the objective function.

More precisely, the One-Class SVM solving the following quadratic optimization problem,

$$\min \left(R^2 + C \sum_{i=1}^{n} \xi_i \right) \tag{9}$$

s.t. $(x_i - c)^T (x_i - c) \leq R^2 + \xi_i, \quad \xi_i \geq 0$, where c and R are the center and radius of the sphere.

The tradeoff between the radius of the hyper-sphere and the number of training samples that it can hold is set by the parameter C. When C is small, the optimization is tend to put more data into the hyper-sphere. When C is larger, we try to squeeze the size of the hyper-sphere.

5.3 Time-Span Penalty Factor

According to our hypothesis, all the samples get from training video are in the period of normal condition, so after time-span filtering, the final results need to approach continuous normality. Under this analysis, we classify the false alarm into faculative false alarm and intrinsically false alarm. We believe that faculative false alarm does not affect the final result in general. So we need to decrease the intrinsically false alarm as much as possible. We present an improved one-class SVM training method to solve this problem. In traditional one-class SVM training, the optimization does not consider the timeline relationship between training samples. We add a punishment term into the optimization problem as follows:

$$\min \left(R^2 + C_1 \sum_{i=1}^{n} \xi_i + C_2 \sum_{\mathrm{j}=1}^{\mathrm{m}} \omega_{\mathrm{j}} \right) \tag{10}$$

where the third term shows the punishment in consequent false alarm in time span, representing the intrinsically false alarm.

6. Depression Angle and Motion Mapping

6.1 Reflecting the Motion Direction on Ground Plane

We need to use motion in ground plane to describe normal or abnormal events. But the motion in ground plane can not be obtained directly. The motion vectors we can get directly are the motion vector in frame. So we need to use motion vector in frame to calculate motion in ground plane. In order to do this, we need to set three hypotheses.

- All the objects concerned are below the camera.

- All the motions of object are horizontal
- Ignore the camera distortion

In reality, surveillance video camera is always set on a high position, that is higher than all the concerned objects in the scene. And secondly, in walking video, over 95% or the motion is horizontal. In scuffle video, limb motion is not always horizontal. But we do not care about whether the limb direction in fight is horizontal or vertical so much. Thirdly, the camera distortion is not very serious in surveillance camera. The error caused by camera distortion is under the magnitude of percentage. This is negligible on the following statistics. Under these three hypotheses, we can get the real motion mapping. This mapping is directed by pixel position in video frame while independent of depth from camera, which is unobtainable by video itself. There are two parameters for different positions, first is depression angle, it can be calculate by vertical viewing angleΓ and pixel position.

$$\gamma = \left(\frac{h - h_0}{H}\right)\Gamma \tag{11}$$

where h is the pixel y-axis position and h_0 is the horizontal line pixel position. H is the video resolution height.

The second is horizontal angle which can be calculate by horizontal viewing angleΘ and pixel position.

$$\theta = \left(\frac{w - \frac{W}{2}}{W}\right)\Theta \tag{12}$$

where w is the pixel x-axis position and W is the video resolution width.

With these parameters, we can find the mapping from motion direction on ground β to motion direction in frame α as

$$\tan\alpha = \tan\left(\beta - \tan^{-1}\left(\frac{\tan\theta}{\cos\gamma}\right)\right) / \sin\gamma \tag{13}$$

$$\tan\alpha = \tan\left(\beta - \tan^{-1}\left(\frac{\tan\left(\frac{w-\frac{W}{2}}{W}\right)\Theta}{\cos\left(\frac{h-h_0}{H}\right)\Gamma}\right)\right) / \sin\left(\frac{h-h_0}{H}\right)\Gamma \tag{14}$$

So we can deduce the inverse mapping from motion direction in frame α to motion direction on ground β as follows.

$$\beta = \tan^{-1}\left(\tan\alpha\sin\left(\frac{h-h_0}{H}\right)\Gamma\right) + \tan^{-1}\left(\frac{\tan\left(\frac{w-\frac{W}{2}}{W}\right)\Theta}{\cos\left(\frac{h-h_0}{H}\right)\Gamma}\right) \tag{15}$$

An schematic graph of reflect result is shown in Fig. 8 and Fig. 9 set out an example of mapping from α to β with some related parameters.

6.2 Depression Angle

In extreme circumstances, when depression angle is 90 degree, the motion direction on ground β to motion direction in frame α is the same. The mapping from α to β does not produce new errors. When depression angle is 0 degree, the mapping can not be found, or we can not get the motion direction on ground β. Generally speaking, when depression angle is larger, the mapping is more accurate, and when depression angle is smaller, the mapping is more unauthentic. We discuss in this section about in which depression angle range the mapping is trustworthy.

Fig. 8 Reflecting the motion direction on ground plane.

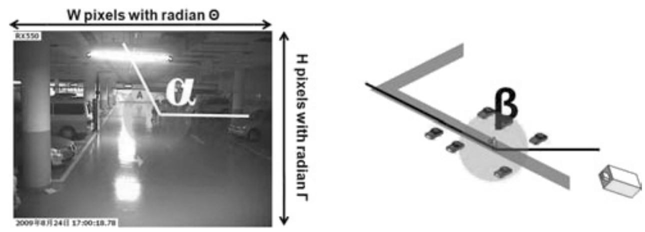


Fig. 9 Mapping from α to β.

We do error analysis on β's calculation, the only independent variable is the motion direction in frame α.

$$\begin{aligned}\Delta\beta &= \frac{d\beta}{d\alpha}\Delta\alpha \\ &= \frac{1}{1 + \left(\tan\alpha\sin\left(\frac{h-h_0}{H}\right)\Gamma\right)^2}\sin\left(\frac{h-h_0}{H}\right)\Gamma\sec^2\alpha\Delta\alpha\end{aligned} \tag{16}$$

It can be used in a different form of expression as

$$\Delta\beta = \frac{1}{\cos^2\alpha + \sin^2\alpha\sin^2\left(\frac{h-h_0}{H}\right)\Gamma}\sin\left(\frac{h-h_0}{H}\right)\Gamma\Delta\alpha \tag{17}$$

We can estimate the mean square error when calculating β from α at different depression angle γ.

$$MSE = \frac{1}{2\pi}\int_0^{2\pi}(\Delta\beta)^2 d\alpha|_\gamma \tag{18}$$

We find that if the depression angle is lower, the error magnification from α to β is larger and harder to use. We use Fig. 10 to estimate the depression angle threshold. The bin difference in motion directional histogram is 22.5 degrees. When the mean square error of β is over half of the difference, we decide the statistics is trustless. From that, we can introduce a lower limit of depression angle of our method as 9 degrees.

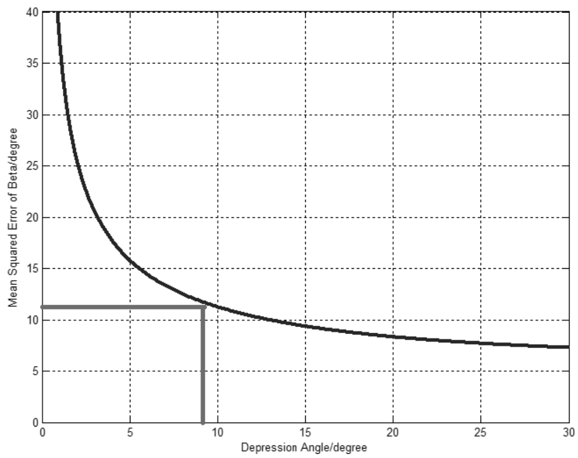


Fig. 10 Mean square error of β under different depression angle.

7. Experiments

7.1 Datasets

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

As shown in Fig. 11, 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 2.5hr duration, events including single-walking, group-walking, talking, meeting and fighting were captured by two camera.

The last dataset we used is CASIA dataset. The dataset's resolution is 320×240 with 25 frames per second and was compressed with the huffyuv code in avi format. We use this dataset as no-training test. We test our algorithm without scene demarcation and training. We used the model that has trained in the underground parking dataset to test this scene. In other words, the model does not use any information from the same dataset. The ability to detect anomaly is from universal character. This experiment shows our model's commonality and universality. After training a normal model, the model can be used in all the other surveillance video with little degradation in the quality.

7.2 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 leaving 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 datasets. A frame number of T=50 is used. In the PETS2004 dataset (Fig. 12), 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 videos with no false alarm in the other 17 normal videos. The only incorrectly rejected

Fig. 11 Example frames of Underground parking dataset.

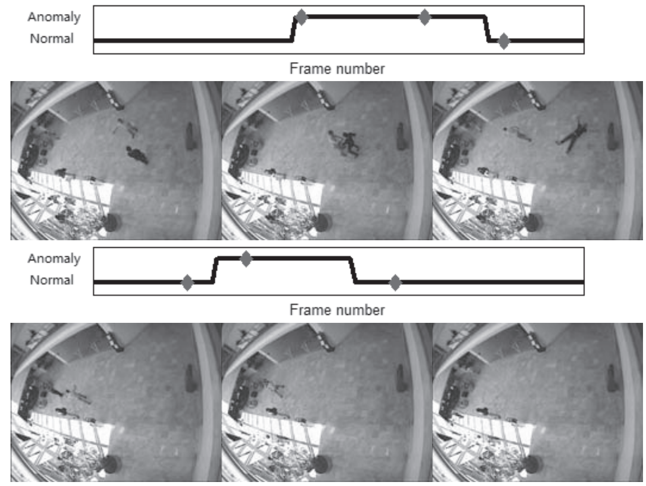


Fig. 12 Detected anomaly periods (The time when black line is at high level is the time of anomaly detected. It is the same for the following figures) and example frames (The time of the dots) in PETS2004 dataset. (Frame 105, 155, 187 in Fight_OneManDown and Frame 139, 164, 224 in Fight_RunAway2) (A short miss-detecting period around Frame187 in Fight_OneManDown)

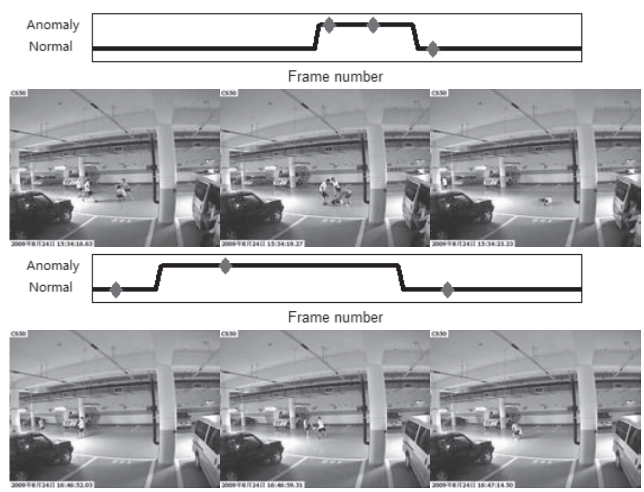


Fig. 13 Detected anomaly periods and example frames (The time of the dots) in Underground Parking dataset A.

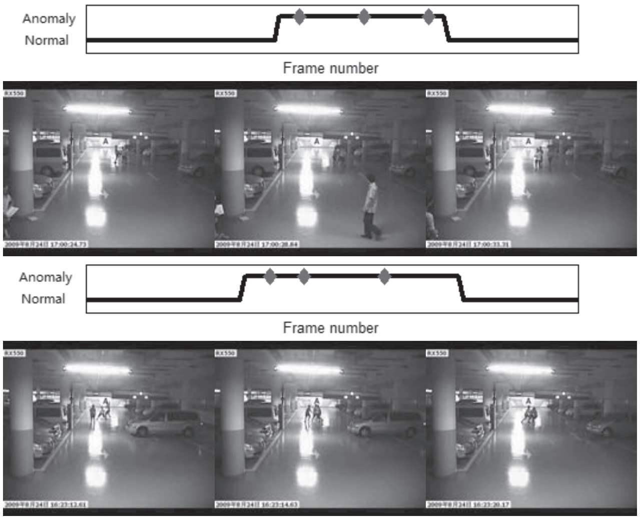


Fig. 14 Detected anomaly periods and example frames (The time of the dots) in Underground Parking dataset B (The algorithm described in Sect. 6.1 is used and has contributed greatly).

Fig. 15 Example frames of precise anomaly detection in CASIA dataset.

abnormal video is *Fight_Chase.mpg* in which the fighting only last for a very short duration and the limb movements are not intensive.

In the Underground Parking dataset (Fig. 13), there are 17 fighting events in the 2.5hr duration. We used 25 minutes of the video without fighting as normal samples. After training the anomaly detector, we test it on the complete 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 algorithm described in Sect. 6.1 plays an important role in Underground Parking dataset B (Fig. 14). In this dataset, the depression angle of camera is lower. The precision rate of abnormal period increases from 47% to 83% after using this algorithm.

In the CASIA test (Fig. 15), we test our algorithm without scene demarcation and training. The model is trained by features in Underground Parking dataset A. All the normal video are detected as normal and over 70% of 'Fight' and

Table 1 Different datasets results.

Results		Detect as Abnormal	Detect as Normal	Precise Rate
PETS2004 Dataset	Normal	0 s	387 s	100%
	Abnormal	17 s	3 s	85%
Underground Dataset A	Normal	180 s	8268 s	98%
	Abnormal	364 s	46 s	89%
Underground Dataset B	Normal	204 s	8110 s	98%
	Abnormal	339 s	71 s	83%
CASIA Dataset	Normal	0 s	1032 s	100%
	Abnormal	100 s	41 s	71%

'Punching a car' video shows an alarm in the detection results. In another word, the false alarm is 0% and the detection rate is over 70%. From this experiment, we prove that the anomaly detector has the universality. The detector is not scene-specific. By using the detector training under one specific scene, we can utilize it on other scenes as well. On one hand, the detector is universal and acceptable for different scene; on the other hand, further retrain for specific scene will improve the performance. The results are summarized in Table 1.

8. Conclusion

In this paper, we have proposed an anomaly detection method in surveillance video. Our algorithm employed a feature of motion vector statistics. The proposed feature is easy to calculate and does not rely on any detection or tracking module, making the system easier to handle event detection in complicated situations. A motion directional PCA was applied for feature down dimension, which is important for the efficiency of subsequent training. One-class SVM was employed to overcome the problem of lacking abnormal data. Experimental results show that most drastic anomaly events are detected by the proposed algorithm with a low false alarm rate. The proposed algorithm do not have the ability to detect calm anomaly, such as person's collapse or hesitation. It should be considered in the following works. In the future, we will also consider anomaly detection based on multi-camera analysis. The results from single-camera will be involved into a broader range to understand.

Acknowledgement

Portions of the research in this paper use the CASIA Action Database collected by Institute of Automation, Chinese Academy of Sciences.[19]

We also acknowledge with thanks that the research was supported by Sony China Research Lab.

References

[1] T. Ding, M. Sznaier, and O.I. Camps, "Fast track matching and event detection," CVPR, pp.1–8, June 2008.
[2] F. Jiang, Y. Wu, and A.K. Katsaggelos, "Abnormal event detection based on trajectory clustering by 2-depth greedy search," Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, pp.2129–2132, March-April 2008.

[3] F. Cupillard, A. Avanzi, F. Bremond, and M. Thonnat, "Video understanding for metro surveillance," Proc. IEEE International Conference on Networking, Sensing and Control, pp.186–191, March 2004.
[4] H. Zhong, J. Shi, and M. Visontai, "Detecting unusual activity in video," CVPR, vol.2, pp.819–826, 2004.
[5] Y. Ke, R. Sukthankar, and M. Hebert, "Efficient visual event detection using volumetric features," IEEE Computer Society, ICCV, pp.166–173, Washington, DC, USA, 2005.
[6] P. Cui, L. Sun, Z. Liu, and S. Yang, "A sequential monte carlo approach to anomaly detection in tracking visual events," CVPR, pp.1–8, June 2007.
[7] W. Lao, J. Han, and P. de With, "Automatic surveillance analyzer using trajectory and body-based modeling," Proc. Digest of Technical Papers International Conference on Consumer Electronics, pp.1–2, Jan. 2009.
[8] Y. Wang, K. Huang, and T. Tan, "Human activity recognition based on r transform," CVPR, pp.1–8, June 2007.
[9] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," CVPR, pp.1446–1453, June 2009.
[10] X. Wang, X. Ma, and E. Grimson, "Unsupervised activity perception by hierarchical bayesian models," CVPR, 2007.
[11] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," CVPR, vol.0, pp.935–942, 2009.
[12] N. Ihaddadene and C. Djeraba, "Real-time crowd motion analysis," ICPR, pp.1–4, Dec. 2008.
[13] 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, July 2008.
[14] D. Tax and R. Duin, "Support vector domain description," Pattern Recognit. Lett., vol.20, pp.1191–1199, 1999.
[15] D. Lin, E. Grimson, and J. Fisher, "Learning visual flows: A lie algebraic approach," CVPR, vol.0, pp.747–754, 2009.
[16] A.K. Jain, Fundamentals of Digital Image Processing, Prentice Hall, Upper Saddle River, NJ, USA, 1989.
[17] S.J. Hanson, G.S. new Brunswick, C. Kulikowski, and N. Japkowicz, "Concept-learning in the absence of counter-examples: An autoassociation-based approach to classification," Technical Report, July 1999.
[18] J. Laaksonen, M. Koskela, and E. Oja, "Picsom: Self-organizing maps for content-based image retrieval," IJCNN '99, pp.2470–2473, July 1999.
[19] CASIA action database, http//www.sinobiometrics.com/

Chang Liu was born in 1984. He received his B.S. degree in department of Electronics Engineering, Tsinghua University, China in 2006. He is currently pursuing the Ph.D. degree at Department of Electronics Engeering, Tsinghua University, China. His research interests are focused on image and video processing, pedestrian detection and tracking, surveillance video analysis, etc.

Guijin Wang was born in 1976. He received the B.S. and Ph.D. degree (with honor) from the department of Electronics Engineering, Tsinghua University, China in 1998, 2003 respectively, all in Signal and Information Processing. From 2003 to 2006, he has been with Sony Information Technologies Laboratories as a researcher. From Oct., 2006, he has been with the department of Electronics Engineering, Tsinghua University, China as an associate professor. He has published over 30 International journal and conference papers, hold several patents. He is the session chair of IEEE CCNC'06, the reviewers for many international journals and conferences. His research interests are focused on wireless multimedia, mesh network, image and video processing, object detection and tracking, online learning, etc.

Wenxin Ning was born in 1985. He received his B.S. degree in department of Electronics Engineering, Tsinghua University, China in 2008. He is currently pursuing the M.S. degree at Department of Electronics Engeering, Tsinghua University, China. His research interests are focused on image and video processing, surveillance video analysis, etc.

Xinggang Lin received B.S. degree in Electronics Engineering, Tsinghua University, China in 1970; M.S. degree in 1986 and Ph.D. degrees in 1982, both in information science, Kyoto University, Japan. He joined the Department of Electronics Engineering at Tsinghua University in 1986 where he has been a full professor since 1990. He received "Great Contribution Award" from Ministry of Science and Technology of China, and "Promotion Awards of Science and Technology" from Beijing Municipality. He was a General Co-chair of the second IEEE Pacific-Rim Conference on Multimedia, an associate editor of IEEE T. on CSVT, and a technical/organizing committee member of many international conferences. He is a fellow of China Institute of Communications, and he published over 140 referred conference and journal papers in diversified research fields.