MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

Detecting the Degree of Anomal in Security Video

Kyoko SUDO*, Tatsuya Osawa, Xiaojun Wu, Kaoru Wakabayashi and Takayuki Yasuno
NTT Cyber Space Laboratories, NTT Corporation

Abstract

We have developed a method that can discriminate anomalous image sequences for more efficiently utilizing security videos. To match the wide popularity of security cameras, the method is independent of the camera setting environment and video contents. We use the spatio-temporal feature obtained by extracting the areas of change from the video. To create the input for the discrimination process, we reduce the dimensionality of the data by PCA. Discrimination is based on a 1-class SVM, which is a non-supervised learning method, and its output is the degree of anomaly of the sequence. In experiments we apply the method to videos obtained by a network camera; the results show the feasibility of indexing anomalous sequences from security videos.

1 Introduction

The number of security video systems is increasing rapidly. Many cameras are now present in public spaces. The problem is that there is a shortage of personnel to check the enormous numbers of the captured videos, so an automatic surveillance method that can raise the efficiency of video checking is required. To reduce the time and expense needed for manual confirmation, one effective approach is to tag sequences as either normal or anomalous.

Identifying the change areas is important in anomalous sequence detection. Change areas can be extracted by background subtraction, however, when the background image cannot be taken because people move continuously across the field of view or because the background itself changes, the background has to be estimated. Also, anomalous sequences contain not only frames in which people keep moving but also frames in which people stay still, so the object extraction method must be able to extract anomalous sequences regardless of the actions of the people captured. For this reason, we use a background estimation method with a renewal equation; this equation controls the influence of the prior background by setting a timing parameter.

After extracting the change areas, we have to discriminate the anomalous samples by training. We take the non-supervised approach, because in many cases of anomaly detection, we don't know what will be anomalous in advance. There are a few methods that take the non-supervised approach in detecting anomalies. One is to make acceptable motion patterns by clustering, and then use the clusters to detect anomalous events[1], [2]. Unfortunately, the structures of clusters are sometimes complicated and not intuitive in discerning the distribution of patterns. Another method is to learn many samples and make a linear subspace[3][4]. Our approach is based on a 1-class SVM, and we use its output as the degree of anomaly of the sequence.

*Address: 1-1 Hikari-no-oka Yokosuka Kanagawa 239-0847, Japan. E-mail: sudo.kyoko@lab.ntt.co.jp

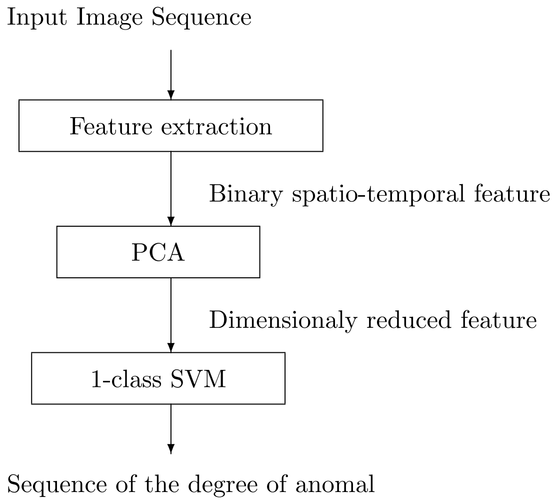


Figure 1: Process of extracting the degree of anomal from the input image sequence.

We have applied the method to videos containing long sequences of up to a few minutes, using the feature produced from over a hundred frames for discrimination[5]. However, in many cases, security videos from real surveillance systems consists of a lot of short video sequences of several minutes, since they are captured when motion is detected by some motion detection function of the camera. We have conducted several experiments in which we applied the method to those kind of videos using a feature extracted from just 15 frames for discrimination. The results show that the degree of anomal extracted by our method allows the addition of reasonable tags and so well reduces the volume of videos that must be reviewed.

2 Feature for anomaly extraction

The algorithm of our method is shown in Figure 1.

Features based on moving areas are effective for detecting anomalies because the existence, or otherwise, of moving objects is important information. To specify the kind of motion, a single frame of a sequence that contains motion is insufficient. For this, we derive a spatio-temporal feature. First, we extract the moving areas in each frame. We then obtain a binary image sequence in which each frame has a foreground value of 1 and background value of 0. To obtain the

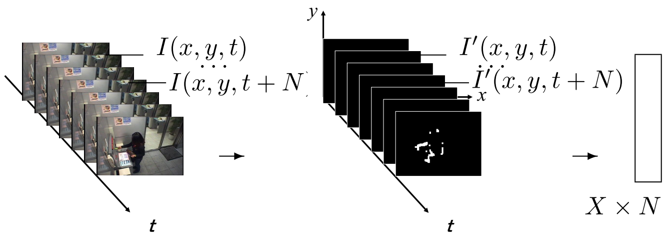


Figure 2: Process of extracting the spatio-temporal feature. X is the number of the pixels in the image. N is the number of frames.

binary image sequence, we create a background model by analyzing the time sequences of each pixel as a mixture distribution[6]; this approach can robustly extract moving objects even if they remain stationary for an appreciable time.

The sequence is divided into sets of a constant number of frames to yield the feature sets. To prevent the problem that the dimensionality of the feature is too high to input it directly into the 1-class SVM module, we compress the dimensionality of the feature by principal component analysis(PCA).

We detail how to obtain the binary image sequence and the spatio-temporal feature below. Let the input pixel at position (x, y) at time t be $X_t (= I(x, y, t))$. The background at time t is modeled by k Gaussian mixture distribution $\nu_{k,t}$. Parameter $M_{k,t}$ is defined as having a value of 1 when X_t lies in the range of the background distribution, and 0, otherwise. The value of weight $w_{k,t}$ of $\nu_{k,t}$ is increased as X_t approaches the center of the kth element distribution. X_t is estimated as the background pixel and output pixel value $X'_t (= I'(x, y, t))$ is set to 0 when $M_{k,t} = 1$. Parameters μ_t and σ_t of distribution $\nu_{k,t-1}$ are updated by adding the value of X_t to $\nu_{k,t-1}$ using equations (2) and (3). In the equation (3), η is a Gaussian function. The speed of updating is controlled by the value of α. X_t is estimated as the foreground pixel and output pixel value $I'(x, y, t)$ is set to 1 when $M_{k,t} = 0$. This yields the binary image sequence $I'(x, y, t)$ $(t = 1, 2, \cdots, T)$.

$$
\begin{aligned}
w_{k,t} &= (1 - \alpha) w_{k,t-1} + \alpha (M_{k,t}) && (1) \\
\mu_t &= (1 - \rho) \mu_{t-1} + \rho X_t && (2) \\
{\sigma_t}^2 &= (1 - \rho) {\sigma_{t-1}}^2 + \rho (X_t - \mu_t)^T (X_t - \mu_t) && (3) \\
\rho &= \alpha \eta (X_t | \mu_k, \sigma_k)
\end{aligned}
$$

After obtaining the binary image sequence, we consider each image as a 1-D vector whose size is $x \times y$ $(= X)$.

To make the following discrimination process efficient, we reduce the size of X by applying PCA to the sequence $I'(x, y, t)$. We use the first pth component, and obtain the principal component feature sequence, $F(t)$ $(t = 1, 2, \cdots)$. After obtaining $F(t)$, we cut it into small sequences. We then create a set of matrixes whose dimension is $p \times N$, p is the size of the dimensionally-reduced feature of the image , and N is the length of each small sequence cut from the whole sequence. We determine the matrix whose components are $F(t - N), F(t - (N - 1)), \cdots, F(t)$ as the feature(Figure 2).

3 Extracting Anomaly Level by 1-class SVM

It is considered that outliers in the feature space are anomalous samples. We treat the distance of the outlier from the region in which most samples lie as indicating the degree of anomaly of the sample. To identify these samples, we use a 1-class SVM, which is a non-supervised learning method. The 1-class SVM maps the outliers in the input space close to the origin of the high dimensional feature space when using Gaussian Kernel $K(x_i, x) = exp(-\frac{||x_i - x||^2}{\sigma^2})$. We use equation4 as the discrimination function. To solve equation5, the super plane discriminates the sample sets such that the rate of ν of all sample sets lie below the origin. Here, ν is set in advance.

$$f(x) = sign(\omega \Phi(x) - \rho) \tag{4}$$

$$
\begin{gathered}
min_{w \in F, \xi \in R^n, \rho \in R} \frac{1}{2} ||\omega||^2 + \frac{1}{\nu n} \Sigma_i \xi_i - \rho \\
\omega \Phi(x_i) \leq \rho - \xi_i, \;\; \xi_i \leq 0
\end{gathered} \tag{5}
$$

Equations (2), (3) are extended by using the Kernel trick for the non-linear case to yield (6), (7).

$$f(x) = sign(\Sigma_i \alpha_i K(x_i, x) - \rho) \tag{6}$$

$$
\begin{gathered}
min_\alpha \frac{1}{2} \Sigma_{i,j} \alpha_i \alpha_j K(x_i, x_j) \\
\begin{aligned}
0 \leq \alpha_i \leq \frac{1}{\nu n}, \;\; i &= 1, \cdots, n \\
\Sigma_{i=1}^n \alpha_i &= 1
\end{aligned}
\end{gathered} \tag{7}
$$

A discrimination axis that maximizes the distance of all samples from the origin is decided by the 1-class SVM by optimizing equation (7). The samples of constant rate ν, which is set in advance, become outliers. In the feature space, as the distances between sample x and all other samples increases, value $\Sigma_i \alpha_i K(x_i, x) - \rho$ in function $f(x)$ in equation (6) becomes smaller. Sample x is considered anomalous when the value of $\Sigma_i \alpha_i K(x_i, x) - \rho$ is negative, and in that case, we use the scalar of $|\Sigma_i \alpha_i K(x_i, x) - \rho| (= |g(x)|)$ as the degree of anomaly. As $|g(x)|$ increases, sample x is considered to be more anomalous. Since the mapping process is non-linear, the size of $g(x)$ does not directly represent the distance between samples in the original feature space. Scholkopf presented experiments on two dimensional feature data. He found that the discriminant boundary changes from the center of the distribution to outside, like a contour line, when ν is increased[7]. Our preliminary experiment using a two dimensional small data set showed that there is an order relation between the size of ν and the degree

Figure 3: A frame of original video.

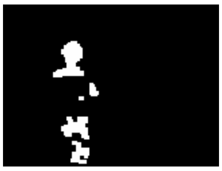


Figure 4: A binary image extracted by background model estimation (extracted from the original image of Figure 3)

of separation between the origin and x. The constant value of σ must be set appropriately. The value of ν indicates what percentage of all samples become outliers. The user sets ν according to the volume of videos that the reviewer wants to examine.

4 Experiment

4.1 Video data

To estimate the performance of the proposed method, we conducted an experiment using a security video captured at an room entrance; a network camera was set about $2m$ above the floor and angled to observe people entering and leaving the room. The camera had a motion detection function, so the videos consisted of many cuts. In order to assess the proposed method, we labeled all video sequences as either normal or anomalous. To eliminate the impact of subjective decisions, we simply define normal sequences as those in which single people enter the room, otherwise anomal. The anomalous sequences showed scenes wherein single or multiple people entered or left the room. The video consisted of jpeg images, each 160×120 pixels, stored on a hard disk at the rate of 15 frames per second with time codes. A sample frame of a cut is shown in Figure 3.

4.2 Feature Extraction

In binary image extraction, the number k of the element distribution of the mixture distribution model was set to 5. The value α of equation (1) was set to 0.01. A binary image sequence was extracted from the original video. A sample of the binary image is shown in Figure 4. Dimension size was reduced to 13 by taking the first 13 principal components of PCA . The sequence was divided into sets of 15 frames to yield the feature sets. The spatio-temporal feature, $(160 \times 120(pixels)) \times 15(frames)$, was extracted starting from every 15th frame; 100 input features were obtained.

The number of input features labeled anomalous was 30.

4.3 Anomaly Extraction

We used the LIBSVM[1] package for SVM processing. We decided the ν value to yield the desired volume of anomalous cuts. The parameters set were $\nu = 0.2$, $\sigma = 0.1$. To decide parameter σ, we conducted an experiment using small subsets of various σ values, ν was held constant, and selected the value that yielded the best performance.

4.4 Experimental Results

The result of sample frames from the sequences discriminated as anomalous and normal are shown in Figure 5. It is considered that any input feature that yielded a negative $g(x)$ value was anomalous. The results show that movement type was well discriminated is either normal or anomal. As a result, all sequences manually tagged as anomalous in section4.1 were present in the 20 highest degree of anomal sequences; the first 20 sequences (according to time code) contained only 20% of the anomal sequences. This shows that sorting sequences by their degree of anomal is very efficient; the alternative is to check all sequences.

5 Conclusion

We proposed a method that can identify anomalous sequences in output of security cameras. It is characterized by the use of a spatio-temporal feature; no heuristics is used. it is based on non-supervised learning via a 1-class SVM and so does not need prior labeled data. We use the discrimination function of the 1-class SVM to identify anomalies. The proposed method was applied to a video captured by a network camera. The results show that the regions with high degrees of anomaly contain the cuts labeled anomalous. We can efficiently find minority movements by sorting sequences by the degree of anomal as indicated by our method. We intend to improve the algorithm so that training is performed incrementally by adding new samples to the original data set.

References

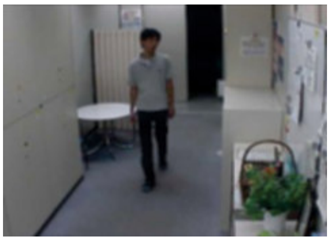
[1] Z. Fu, W. Hu and T. Tan: "Similarity Based Vehicle Trajectory Clustering and Anomaly Detection", Proc. IEEE International Conference on Image Processing, ICIP2005.

[2] A. Mecocci and M. Pannozzo: "A completely autonomous system that learns anomalous movements in advanced video surveillance applications", Proc. IEEE International Conference on Pattern Recognition, ICPR2005.

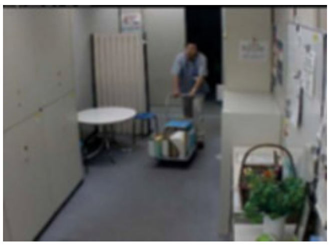
[3] T. Nanri and N. Otsu: "Unsupervised Abnormality Detection in Video Surveillance", Proc. IAPR Conference on Machine Vision Application, pp. 574–577, 2005.

[1] http://www.csie.ntu.edu.tw/ cjlin/libsvm

(a)



(b)



(c)

Figure 5: Examples of normal and anomalous sequences. Sequences in which a single person is entering the room are considered normal, all others are anomal.(a) shows the anomalous sequences detected as $g(x) < 0$. $|g(x)| = 0.23$, $|g(x)| = 0.16$, $|g(x)| = 0.14$, from top to bottom. Those images show the sequences in which a man is leaving the room, a man is going out with a cart, two men are walking in opposite directions, each. (b)sequence labeled normal: $g(x) > 0$ ($g(x) = 0.11$). The image shows a sequence in which a man is leaving the room. This type of sequence is the majority of the input. (c)sequence labeled nomal but detected as anomal: $|g(x)| = 0.004$.

[4] E. Andrade, S. Blunsden, and R. Fisher: “Modelling Crowd Scenes for Event Detection”, Proc. ICPR2006.

[5] K. Sudo, T. Osawa, K. Wakabayashi and T. Yasuno: “Detecting the Degree of Anomaly in Security Videos by using a Spatio-Temporal Feature of Change,”in Proceedings of the International Conference on Signal and Image Processing, 2006.

[6] Stauffer, Grimson：“Adaptive background mixture models for real-time tracking”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol.2, pp.246–252,1999.

[7] B. Scholkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola and R. C. Williamson: “Estimating the support of a high-dimensional distribution”, Neural Computation **13**, pp.1443–1471, 2001.