Detecting the Degree of Anomaly in Security Videos

Kyoko Sudo†, Tatsuya Osawa, Kaoru Wakabayashi, and Hideki Koike

Abstract

We introduce a method of detecting video scenes that require attention and presenting them in order of significance. We use a statistical method to detect a pattern that differs from the *regular* pattern and extract the distance between them as an anomaly. If this method is applied to monitoring systems, it should greatly reduce the cost of checking a huge amount of video data.

1. Introduction

With the rapid increase in the number of video security systems, one new problem is how to manage the enormous amount of video data being captured and stored over networks. The cost of reviewing and checking security videos is extremely high, so an automatic surveillance method that can efficiently check videos is required in order to reduce the time and expense of manual confirmation. One idea is anomaly detection, that is, to discriminate video images from the viewpoint of whether or not they look like ordinary scenes [1]. After this discrimination, we can tag sequences as either normal or anomalous. Tagging reduces the volume of videos that must be reviewed.

To discriminate the anomalous samples by training, one existing method finds anomalous sequences by using prior knowledge or by learning normal data [2]. If we define the steady state as some feature pattern, we can discriminate an anomalous feature pattern by comparing it against the steady-state feature pattern. However, in many cases of anomaly detection, we do not know in advance what might be anomalous. Accordingly, we decided to take the nonsupervised approach.

† NTT Cyber Space Laboratories
Yokosuka-shi, 239-0847 Japan
Email: Sudo.kyoko@lab.ntt.co.jp

2. Our approach

2.1 Features for anomaly detection

Features based on areas of movement are effective for detecting anomalies because the presence or absence of moving objects is important information. One major feature extraction method for detecting anomalies is extracting the tracks of moving objects [3]–[5], which is effective when the tracks are labeled as normal or anomalous. Another method uses a local spatio-temporal feature that can discriminate anomalous scenes based on the types of action [6]. In our research, we use a spatio-temporal feature that can extract information from sufficiently long periods to discriminate anomalous scenes based on a sequence of human movement.

First, we extract the areas containing movement in each frame. Our approach is to estimate the distribution of image pixel values as a probabilistic model [7] and then subtract the estimated background. We then obtain a binary image sequence in which each frame has a foreground value of 1 and background value of 0. The sequence is divided into sets of a constant number of frames to yield the feature sets. One problem is that the dimensionality of the feature is too high to allow the feature to be input directly into the statistical discrimination module. Our solution is reduce the dimensionality of the feature by principal component analysis (PCA). We use the dimensionally reduced feature by taking a small number of principal components—those whose contribution rates are

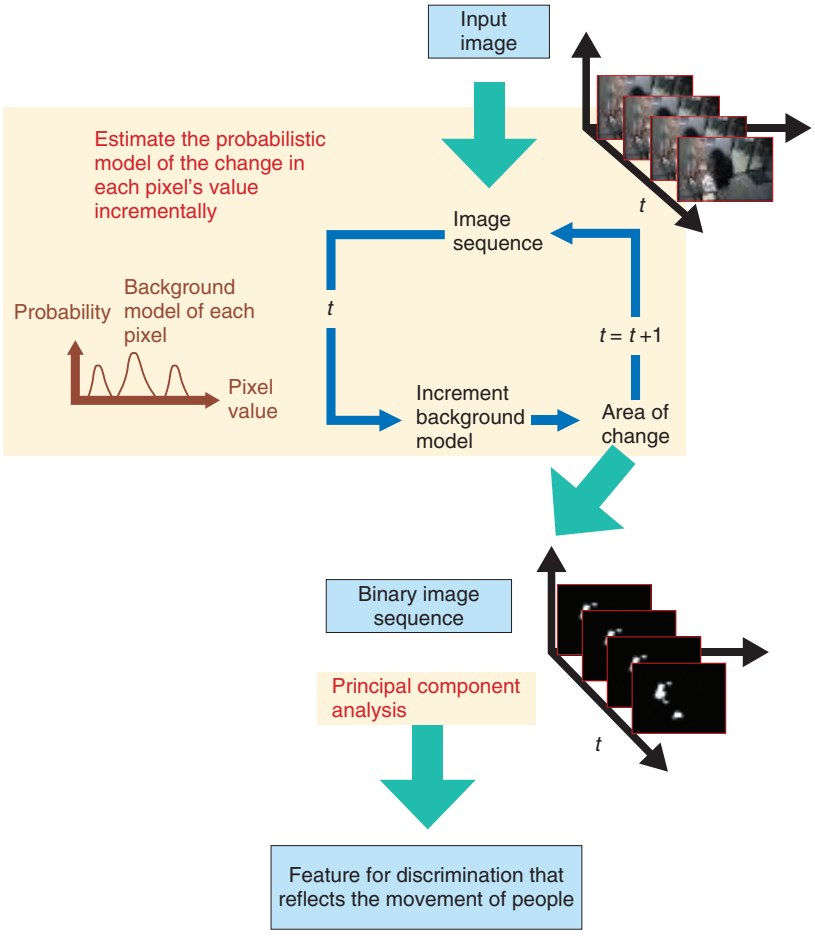

Fig. 1. Process of obtaining the spatio-temporal feature for discriminating anomalous movements.

adequate. We examine the first n principal components (ranked in order of decreasing eigenvalue), where n is defined as the number that makes the ratio of the sum of the first 1-n eigenvalues divided by the sum of all eigenvalues of the variance-covariance matrix equal to 0.9.

After obtaining the binary image sequence, we consider each image as a one-dimensional 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, …). After obtaining $F(t)$, we cut it into small sequences. We then create a set of matrices whose dimension is $p \times N$, where 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 $F(t\text{-}N)$, $F(t\text{-}(N\text{-}1))$, …, $F(t)$ as the feature. The algorithm is shown in **Fig. 1**.

This feature enables even very small movements to be robustly obtained and can reflect timing information about movements. For example, when the movement of a person traces a different track from those of other persons, the area of the silhouette in the spatio-temporal space is different from those of others and a different feature is obtained. This suggests that suspicious movements can be detected. Figure 1 also shows the process for obtaining the spatio-temporal feature for discriminating anomalous movements. After the binary images have been extracted by subtracting the background, whose model is estimated incrementally, the images are dimensionally reduced by PCA. The sequences of principal components are used as the feature for discrimination.

2.2 Detecting the degree of anomaly using a 1-class SVM

Outliers in feature space are treated as anomalous samples. There are several algorithms that can detect outliers in sample distributions. One defines a sample as an outlier if it is not contained within any pre-

learned classes [2]; others are based on clustering [3]–[5], based on estimating a probability function or using the subspace method [6]. In our approach, we use a 1-class support vector machine (SVM) [8], which is a nonsupervised outlier detection method. By optimizing the nonlinear evaluation function, it determines the axis on which some samples are discriminated as outliers. On this axis, the degree of anomaly can be quantitatively extracted as the distance from the outlier to the major distribution containing most of the samples.

The 1-class SVM maps the outliers in the input space close to the origin of the high-dimensional feature space when using a Gaussian kernel

$$K(x_i, x) = exp\left(-\frac{||x_i - x||^2}{\sigma^2}\right).$$

We use Eq. 1 as the discrimination function. To solve Eq. 2, the super plane discriminates the sample sets such that the rate ν of all sample sets lies below the origin. Here, ν is set in advance.

$$f(x) = sign(\omega \Phi(x) - \rho), \quad (1)$$

$$\min_{w \in F,\ \xi \in R^n,\ \rho \in R} \frac{1}{2}||\omega||^2 + \frac{1}{\nu n}\sum_i \xi_i - \rho \quad (2)$$

$$\omega \Phi(x_i) \leq \rho - \xi_i,\ \xi_i \leq 0.$$

Equations (1) and (2) are extended by using the kernel trick for the nonlinear case to yield Eqs. (3) and (4).

$$f(x) = sign(\sum_i \alpha_i K(x_i, x) - \rho), \quad (3)$$

$$\min_\alpha \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) \quad (4)$$

$$0 \leq \alpha_i \leq \frac{1}{\nu n},\ i = 1, \ldots, n$$

$$\sum_{i=1}^n \alpha_i = 1.$$

A discrimination axis that maximizes the distance of all samples from the origin is determined by the 1-class SVM by optimizing Eq. (4). The samples with constant rate ν, which is set in advance, become outliers. In the feature space, as the distances between sample x and all other samples increase, the value $\sum_i \alpha_i K(x_i, x) - \rho$ in function $f(x)$ in Eq. (3) becomes smaller. Sample x is considered anomalous if the value of $\sum_i \alpha_i K(x_i, x) - \rho$ is negative, and in that case, we use the scalar of $|\sum_i \alpha_i K(x_i, x) - \rho| (= |f(x)|)$ as the degree of anomaly. As $|f(x)|$ increases, sample x is considered to become more anomalous. Since the mapping process is nonlinear, the size of $f(x)$ does not directly represent the distance between samples in the original feature space. Scholkopf et al. presented experiments on two-dimensional feature data. They found that the discriminant boundary changed from the center of the distribution to outside, like a contour line, when ν was increased [8]. Our preliminary experiment using a small two-dimensional data set showed that there is an order relation between the size of ν and the degree of separation between the origin and x. The constant value σ must be set appropriately. The value of ν indicates what percentage of all samples are outliers. The user sets ν according to the amount of video data that the reviewer wants to examine. The discrimination process using the 1-class SVM is shown in **Fig. 2**.

This algorithm is assumed to be valid when the majority of the samples are composed of sequences of the regular state. For this reason, this discrimination is applicable to monitoring videos in which there is a state that is considered regular. Such environments include the exits/entrances of offices and bank ATM (automated teller machine) areas, where only a small number of people occupy the field of view, one at a time, and the tracks of the people are generally fixed.

3. Experiment

3.1 Data and conditions

To estimate the performance of our method, we conducted an experiment using a security video captured by reproducing a typical setup of a bank ATM. A CCD (charge-coupled device) camera was set about 2 m above the floor and angled to observe people using the ATM. To assess our method, we labeled all the video sequences as either normal or anomalous. These cuts, 40 normal sequences and 10 anomalous sequences, were recorded separately and then merged to yield a 30-minute video. Normal cuts contained sequences in which people stepped up to the cash dispenser, withdrew or deposited money, and then moved away. The cuts containing anomalous sequences showed scenes such as someone removing transaction receipts from the wastebasket. The video was recorded on digital video tape at the rate of 30 frames per second. After compositing, each frame was converted into a JPEG image of 160 × 120 pixels.

The number of dimensions was reduced to 20 by taking the first 20 principal components of PCA. The spatio-temporal feature, 20 × 500 (frames), was

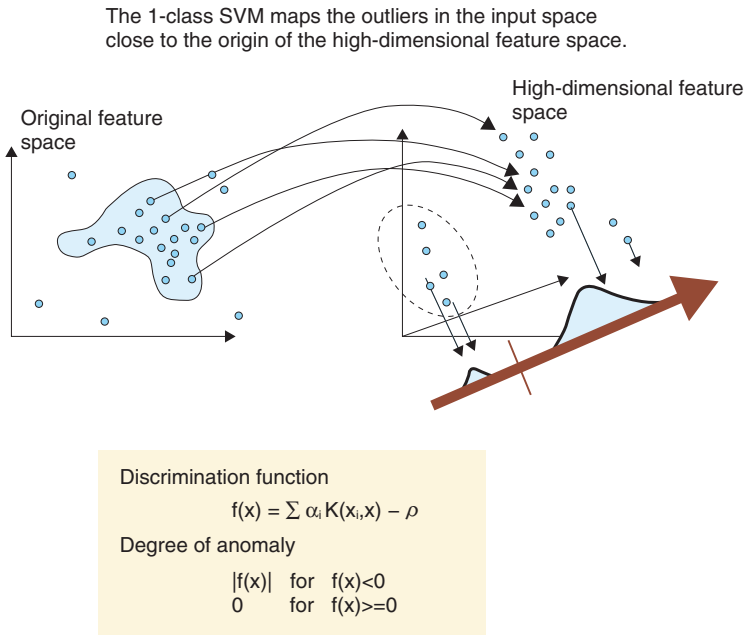

Fig. 2. Discrimination process using 1-class SVM. Samples yielding | $f(x)$ <0 | are identified as anomalous.

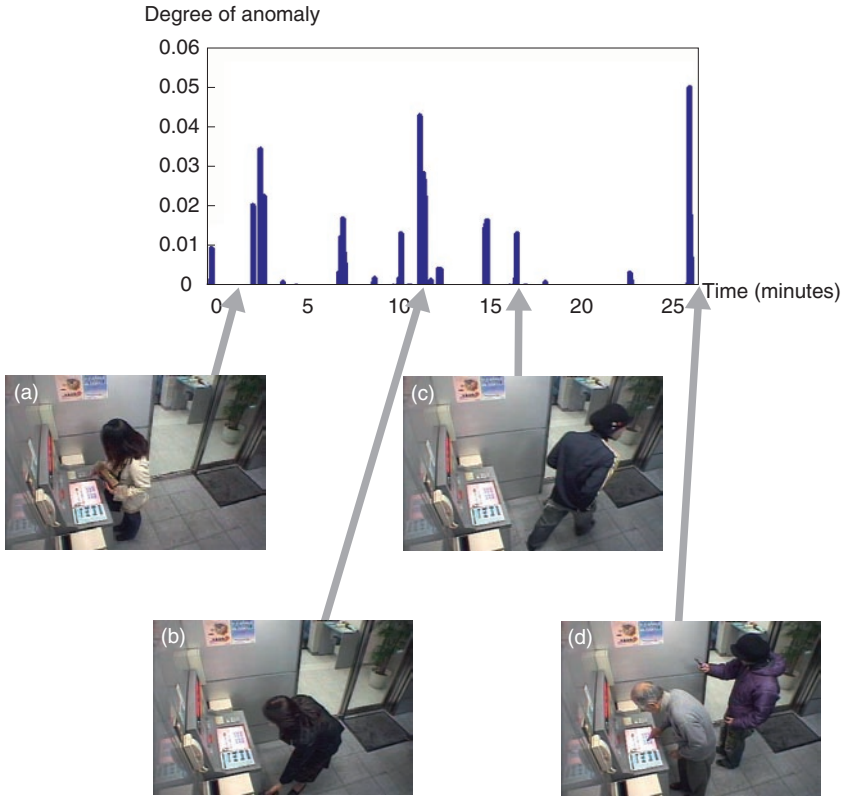

Fig. 3. Degree of anomaly and examples of scenes. (a) Example of a normal scene. (b) Example of a normal scene discriminated as anomalous. (c) Example of an anomalous scene: a man is looking around restlessly. (d) Example of an anomalous scene. A man is taking pictures of the screen on which the person using the ATM is inputting his personal identification number (PIN).

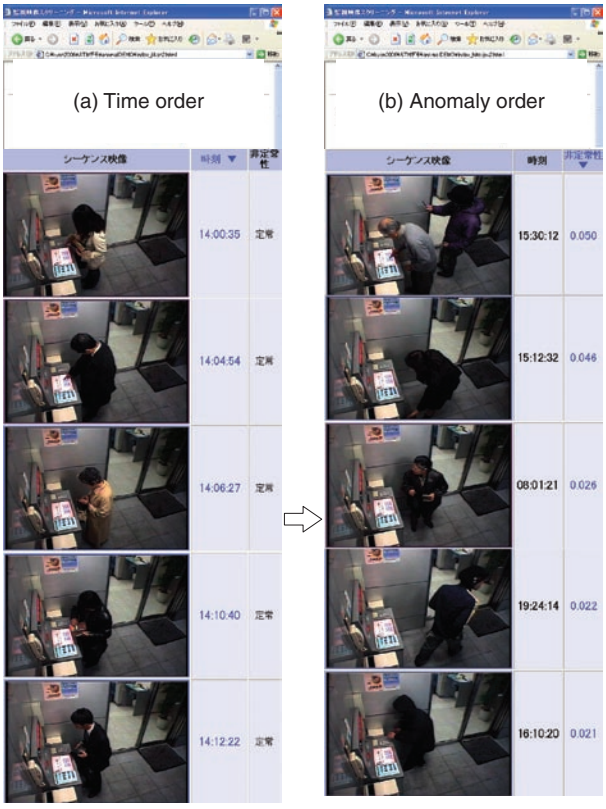

Fig. 4. Results viewer. Image sequences in time order (left) can be sorted by the degree of anomaly (right), which helps manual confirmation.

extracted starting from every 15th frame; 3600 input features were obtained from the 54,000 frames of the 30-minute video. The number of input features that contained anomalous sequences was 150. Those features were labeled as anomalous.

In SVM processing, we set ν to yield the desired volume of anomalous cuts. The parameters were set to $\nu = 0.05$ and $\sigma = 0.01$. To decide σ, we conducted an experiment using small subsets of various σ values, with ν held constant and selected the value that yielded the best performance.

3.2 Experimental results

A graph of the change in degree of anomaly extracted by the 1-class SVM is shown in **Fig. 3**. Any input feature that yielded a negative $f(x)$ value was treated as anomalous. The system assigned negative $f(x)$ values to 124 features, 98 of which were among the 150 features manually labeled as anomalous. As a result, 8 sequences out of the 10 anomalous sequences were detected. Examples of frames from the sequences discriminated as anomalous and normal are also shown in Fig. 3. Sequence (a) manually tagged as normal and sequences (b)–(d) manually tagged as anomalous were extracted. The results show that the movements considered normal and anomalous were well discriminated by our method. In sequence (b), a woman was picking her luggage up from the floor. This was tagged as normal but extracted as anomalous. This is understandable because other people did not perform this action, so it was definitely irregular.

A screenshot of a function for sorting the time-order sequences by anomaly order is shown in **Fig. 4**. This helps people to check videos.

4. Conclusion

We are developing a method that can identify anomalous sequences in security videos. One of its characteristics is that it uses a spatio-temporal feature; no heuristics are used. Another is that it is based on nonsupervised learning using a 1-class SVM, so it does not need prior labeling of data. We use the discrimination function of the 1-class SVM to identify anomalies. Our method was applied to a staged video showing a bank ATM. The video contained a mixture of known normal/anomalous cuts, and the change in the degree of anomaly over time was calculated. The results show that the sequences discriminated as anomalous with high degrees of anomaly contained cuts labeled as anomalous. These results indicate that the degree of anomaly derived by our method closely matches human intuition. Future work includes conducting more extensive trials to discover the limits of this method. We intend to improve the algorithm so that training is performed incrementally because the cost of retraining the system by adding new samples to the original data set is too high.

References

[1] W. Hu, T. Tan, L. Wang, and S. Maybank, "A Survey on Visual Surveillance of Object Motion and Behaviors," Proc. of IEEE Trans. on Systems, Man and Cybernetics, Part C, Vol. 34, No. 3, pp. 334–352, 2004.
[2] M. Gregorio, "The Agent WiSARD Approach to Intelligent Active Video Surveillance System," Proc. of IAPR International Conference on Machine Vision Application, MVA2007, 2007.
[3] S. J. McKenna and H. Nait-Charif, "Learning Spatial Context from Tracking using Penalised Likelihoods," Proc. of IEEE International Conference on Pattern Recognition, ICPR2004, Vol. 4, pp. 138–141, 2004.
[4] Z. Fu, W. Hu, and T. Tan, "Similarity Based Vehicle Trajectory Clustering and Anomaly Detection," Proc. of IEEE International Conference on Image Processing, ICIP2005, Vol. 2, ll-602-5, 2005.
[5] A. Mecocci and M. Pannozzo, "A completely autonomous system that learns anomalous movements in advanced video surveillance applications," Proc. of IEEE International Conference on Pattern Recognition, ICPR2005.
[6] T. Nanri and N. Otsu, "Unsupervised Abnormality Detection in Video

Surveillance," Proc. of IAPR Conference on Machine Vision Application, pp. 574–577, 2005.
[7] C. Stauffer and W. E. L. 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.
[8] 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.

Kyoko Sudo
Research Engineer, Visual Media Communications Project, NTT Cyber Space Laboratories.
She received the B.E. and M.E. degrees in mathematical engineering and information physics and the Ph.D. degree in information physics and computing from Tokyo University, Tokyo, in 1991, 1993, and 2007, respectively. Since joining NTT Laboratories in 1993, she has been engaged in research on image processing and pattern recognition. She is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan.

Tatsuya Osawa
Research Engineer, Visual Media Communications Project, NTT Cyber Space Laboratories.
He received the B.S. degree in physics and the M.E. degree in energy science from Tokyo Institute of Technology, Tokyo, in 2002 and 2004, respectively. Since joining NTT Laboratories in 2004, he has been engaged in research on computer vision. The current focus of his research is on human behavior recognition with distributed cameras. He has received several awards, including the Best Paper Award in ICPR from the International Association for Pattern Recognition, the Funai Best Paper Award from the Funai Foundation for Information Technology, and the Best Paper Award in IVCNZ from the National Group for Image and Vision Computing in New Zealand. He is currently pursuing a Ph.D. degree at Tokyo Institute of Technology. He is a member of IEEE and IEICE.

Kaoru Wakabayashi
Senior Research Engineer, Visual Media Communications Project, NTT Cyber Space Laboratories.
He received the B.E. degree in electro-communications from the University of Electro-Communications, Tokyo, in 1982 and the Ph.D. degree in electronic engineering from the University of Tokyo, Tokyo, in 1999. Since joining Nippon Telegraph and Telephone Public Corporation (now NTT) in 1982, he has been engaged in research on facsimile communications networks, binary image processing, map information processing, cognitive mapping and understanding, and visual monitoring systems. He received the 1993 NTT President's Award, the 1998 AM/FM International Japan Best Speaker Award, the 2006 ICPR Best Paper Award, the 2006 Funai Best Paper Award, and the 2006 IVCNZ Best Paper Award. He is a member of IEICE and the Information Processing Society of Japan.

Hideki Koike
Senior Research Engineer, Supervisor, Group Leader, Visual Media Communications Project, NTT Cyber Space Laboratories.
He received the M.S. degree in mathematics from Tohoku University, Miyagi, in 1985. He joined NTT Labs. in 1985 and engaged in research on image processing. He was transferred to NTT COMWARE in 2001 and engaged in research on RFID. He moved to NTT Cyber Space Labs. in 2007 and is engaged in research on computer vision. He is a member of IEICE.