MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
14-2

Unsupervised Abnormality Detection in Video Surveillance

Takuya Nanri [†]
Nobuyuki Otsu [††] [†]

[†] Graduate School of
Information Science and Technology
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku,
Tokyo 113-8654, JAPAN

[††] National Institute of
Advanced Industrial Science and Technology
(AIST)
1-1-1 Umezono, Tsukuba,
Ibaraki 305-8568, JAPAN

{nanri, otsu}@ isi.imi.i.u-tokyo.ac.jp

Abstract

The detection of abnormal movements is an important problem in video surveillance applications. We propose an unsupervised method for abnormal movement detection in scenes containing multiple persons. Our method uses cubic higher-order local auto-correlation (CHLAC) to extract movement features. We show that the additive property of CHLAC in combination with a linear subspace method is well suited to simplify the learning of normal movements and to detect abnormal movements even in scenes containing multiple persons. One particular advantage of this method is that it does not necessitate the object segmentation and tracking and also any prior knowledge about objects. Some experimental results are shown to exhibit the validity of the method.

1 Introduction

Today, so many surveillance cameras are located all over the place for security purposes. However, it takes too much labor for human to monitor all the data. If only abnormal movements are automatically screened, it saves a lot of labor. Thus, the detection of abnormal movements is a crucial and urgent issue in video surveillance applications.

Assuming surveillance cameras are set up in public places, we must consider that there are simultaneously multiple persons in scenes. Usual strategy in human motion recognition necessitates the segmentation and tracking of each person as a preprocessing [1, 2], and this naturally requires heavy computational load in proportion to the number of persons. Moreover, the recognition accuracy depends on the accuracy of the segmentation and tracking method. In another approach of event-based video analysis [10, 11], it is hard to deal with multiple persons' moves themselves since the extracted features include not only movement features of objects but also a position in an image

In this paper, we adopt CHLAC (cubic HLAC) features [4] for feature extraction to deal with multiple persons. The CHLAC features can extract features of multiple persons' moves without the segmentation or tracking of each person, and computational cost is constant regardless of the number of persons. In many approaches using movement features, prior knowledge about human being is assumed [8, 9], but CHLAC features need no prior knowledge about objects.

For detecting abnormal movements, systems need to recognize those. However it is almost impossible to learn all the examples of rare abnormal (unusual) movements in advance. Nevertheless, we can define abnormal movements as *not being* normal (usual) movements that are often happening in front of a camera. Therefore, abnormal movements can be detected by learning normal movements statistically. In such statistical approaches [7, 11], multiple persons' moves cannot be dealt with directly.

We propose a scheme of combining CHLAC features as movement features and a linear subspace method for learning normal movements. Due to the integral property, CHLAC has a additive property for domain. Therefore, all the normal movements are included in the subspace of normal movements even for scenes containing multiple persons' moves, and only abnormal movements depart from the subspace. It is noticed that normal movements are easily learnt unsupervisedly and the system can construct the subspace of normal movements in an adaptive way.

Further, the subspace of normal movements could incrementally be learnt and constructed in real time. As such online learning, we propose and compare the method solving eigenvalue problem and the incremental method approximating eigenvectors without solving eigenvalue problem (CCIPCA) [3].

2 Present Method

2.1 Preprocessing

We assume a stationary video camera and to extract moving objects from time-differential images. For eliminating noise, we binarize the time-differential images. The threshold at each image is decided by using the discriminant and least squares threshold selection method [6]. **Figure 1** shows an example of image preprocessing.

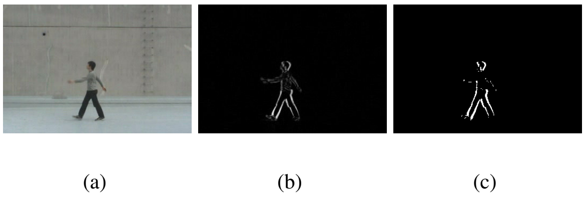


Figure 1: Preprocessing of video frames. (a) original image. (b) time-difference image. (c) binary image.

2.2 CHLAC features

As movement features from time-series binary images, we employ cubic higher-order local auto-correlation (CHLAC) features [4]. CHLAC features for three-dimensional (x, y, t) data are the expansion of higher-order local auto-correlation (HLAC) features [5] for two-dimensional (x, y) data. Each component of CHLAC features is formulated by the following equation:

$$\begin{aligned} & x_f^{(N)}(\boldsymbol{a}_1, \dots, \boldsymbol{a}_N) \\ & \quad \triangleq \int_{W \times H \times T} f(\boldsymbol{r}) f(\boldsymbol{r} + \boldsymbol{a}_1) \dots f(\boldsymbol{r} + \boldsymbol{a}_N) d\boldsymbol{r} \end{aligned} \tag{1}$$

where f is a time-series image, and a variable $\boldsymbol{r}$ and local displacement vectors $\boldsymbol{a}_i \quad (i = 1, \dots, N)$ are three-dimensional vectors in an image data, whose coordinates are *x*-*y* and time. W denotes the width of an image, H the height of an image, and T the range of time.

The number N represents the order of CHLAC, and this paper adopts $N = 0, 1, 2$. Suppose $N = 0$, then this feature represents the number of pixel whose value is 1 in the case of binary images. Suppose $N = 1$, the number of independent displacement vectors is 13. For $N = 2$, it is 237. **Figure 2** shows an example of the second order displacement vector patterns of CHLAC.

Hence, CHLAC feature vectors have $1 + 13 + 237 = 251$ dimensions for the case of binary images.

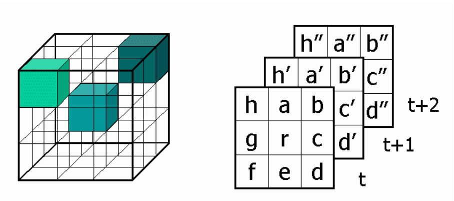
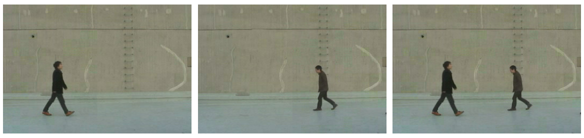


Figure 2: Example of displacement pattern of CHLAC. (h r' b")

CHLAC features have important characteristics, *position invariance* and *additive property*. These characteristics are derived from being integral features. Position invariance means that a feature vector is invariant regardless of location (*x*-*y*-*t*) of a moving object in a data. Additive property is shown in **Figure 3**. Owing to these properties, our method does not require the object segmentation nor tracking, and the computational cost is constant regardless of the number of persons.



(a) f_A (b) f_B (c) $f_A + f_B$

Figure 3: Additive property of CHLAC: The feature vector of (c) is the sum of individuals of (a) and (b), viz. $f_A + f_B$.

2.3 Linear subspace

We use a linear subspace method for the detection of abnormal movements. The reason is that the combination of the additive property of CHLAC features and the linearity of a linear subspace method has a desirable attribute for the detection of abnormal movements. Namely, if the linear subspace of normal movements is constructed, all the normal movement features belong to the subspace and only abnormal movement features depart from the subspace even in the case of multiple persons in a scene. Therefore, we can easily detect abnormal movements by measuring the distance between an input feature vector and the subspace of normal movement features.

At first, we use PCA (Principal Component Analysis) and find the eigenvectors to construct the subspace of normal movements S_N. The eigenvectors $U = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_M], \quad \boldsymbol{u}_i \in V^M (i = 1, \dots, M)$ are calculated by solving the following eigenvalue problem:

$$R_X U = U \Lambda, \qquad R_X \triangleq \mathop{\mathrm{E}}_{i=1}^{n} \{\boldsymbol{x}_i \boldsymbol{x}_i^T\}$$

where $\boldsymbol{x}_i \quad (i = 1, \dots, n) \quad \in V^M$ are M-dimensional feature vectors, and $\Lambda = diag(\lambda_1, \dots, \lambda_M)$ is the eigenvalue matrix. If λ_i are in decreasing order, the contribution rate η_K is represented as

$$\eta_K \triangleq \frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{M} \lambda_i}$$

We adopt the first K eigenvectors for S_N , where K is the smallest number under $\eta_K \geq 0.99$, for example.

The projection operator (projector) onto S_N is given by $P = U_K U_K^T$, where $U_K = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_K]$. Then, the projector onto ortho-complement subspace to S_N is given by $P_\perp = I_M - P$. The distance $d_\perp$ between an input feature vector $\boldsymbol{x}$ and the subspace of normal movements S_N is formulated as

$$\begin{aligned} d_\perp^2 &= \|P_\perp \boldsymbol{x}\|^2 \\ &= \boldsymbol{x}^T (I_M - U_K U_K^T) \boldsymbol{x} \end{aligned}$$

In this paper, we define $d_\perp$ as the *abnormality value*.

The following mathematical expression explains why only abnormal movements are detected by using the projector $P_\perp$. We also illustrate this in **Figure 4**.

$$\begin{aligned} \text{suppose} \quad & \boldsymbol{x} = \boldsymbol{x}_1^{(N)} + \cdots + \boldsymbol{x}_n^{(N)} + \boldsymbol{x}^{(A)} \\ & \quad (N : normal, \quad A : abnormal) \\ \text{then} \quad \|P_\perp \boldsymbol{x}\| &= \|P_\perp (\boldsymbol{x}_1^{(N)} + \cdots + \boldsymbol{x}_n^{(N)} + \boldsymbol{x}^{(A)})\| \\ &= \|P_\perp (\boldsymbol{x}_1^{(N)} + \cdots + \boldsymbol{x}_n^{(N)}) + P_\perp \boldsymbol{x}^{(A)}\| \\ &= \|0 + P_\perp \boldsymbol{x}^{(A)}\| > 0 \end{aligned}$$

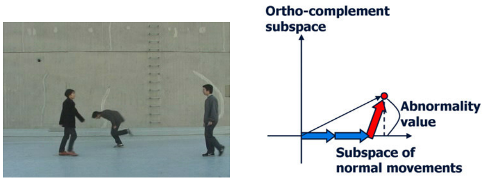


Figure 4: Additive property of CHLAC and linear subspace. (two normal movements (walk) and one abnormal movement (tumble))

2.4 Online learning

Alternatively, the subspace of normal movements can be constructed incrementally for learning normal movements online and in real time and for dealing with a lot of data in real-world applications. We adopt two methods for incremental learning. One is the method which solves eigenvalue problem every step with updating auto-correlation matrix, and the other is the method which approximately calculates eigenvectors without solving the eigenvalue problem.

The former is accurate, but the computational cost is large since the eigenvalue problem is solved every step. The latter is the method using candid covariance-free incremental principal component analysis (CCIPCA) [3]. It is a fast algorithm, since this method does not need to solve the eigenvalue problem. The first eigenvector and the first eigenvalue are updated as follows:

$$\boldsymbol{v}_{n+1} = \frac{n}{n+1}\boldsymbol{v}_n + \frac{1}{n+1}\boldsymbol{x}_{n+1}\boldsymbol{x}_{n+1}^T\frac{\boldsymbol{v}_n}{\|\boldsymbol{v}_n\|}$$

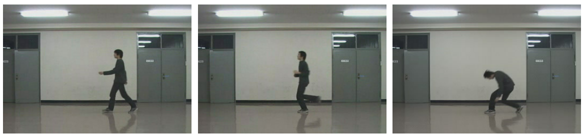
where the first eigenvector is represented as $\boldsymbol{v}_n/\|\boldsymbol{v}_n\|$ and the first eigenvalue as $\|\boldsymbol{v}_n\|$. It is proved that these converge to the true eigenvector and the true eigenvalue as $n \to \infty$. The n-th eigenvector and the n-th eigenvalue are obtained as well.

3 Experiments

We experimented for the case one person appear in a frame and the case multiple persons appear in a frame. In the first two experiments, learning phase was separated from testing phase (batch learning). In the learning phase, the subspace of normal movements was constructed by using only normal movement data, and in the testing phase we compared the distance from the subspace of normal movements. In the last experiment, the subspace of normal movements was learnt by online method without dividing the phase.

3.1 Batch learning

At first we experimented for the case one person appear in a frame. In the learning phase, we adopted “walking” as normal movements, and the learning data contained six objects (persons) that walk rightward or leftward. The testing data contained another person’s walking, running, and tumbling. **Figure 5** shows examples of the testing data.



(a) walk (b) run (c) tumble

Figure 5: Examples for recognition.

Figure 6 shows the distance between an input feature vector and the subspace of normal movements, i.e. abnormality value, at each frame of walking, running and tumbling movement. The distance of normal movement, walking movement, was small, while the distance of running and tumbling movement was both large, so that only abnormal movements was successfully detected. The dimension of the subspace of normal movements was 12 dimensions out of 251 dimensions.

Secondly, we experimented for the case multiple persons appear in a frame by batch learning. In our data, there were up to three persons in a frame simultaneously. In the learning phase, video sequences contained three persons’ walking, and in the testing phase one person out of three walking persons tumbled (**Figure 7**)

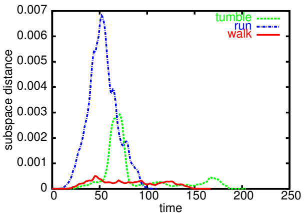
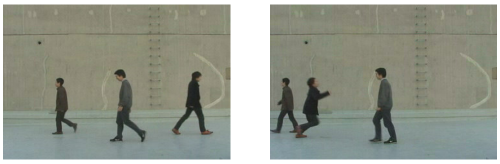


Figure 6: Abnormality value. (one person)



(a) three persons’ walking

(b) one person’s tumbling

Figure 7: Examples of multiple persons’ moves.

Figure 8 shows the abnormality value, i.e. , the distance between an input feature vector and the subspace of normal movements for the data containing only normal movements and for the data containing an abnormal movement. As a result, a tumbling movement in scenes containing two persons’ walking was successfully detected as an abnormal movement. The dimension of the subspace of normal movements was five dimensions out of 251 dimensions.

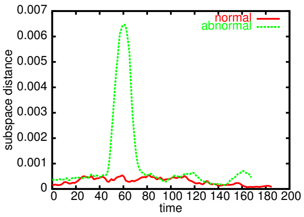


Figure 8: Abnormality value. (batch learning)

In this experiment, a person on the edge of an image and occlusions between persons were not exactly normal movements but a part of normal movements. Nonetheless, experiments showed the features on such occasions hardly had large distances from the subspace. This means that the features on that occasions were also learnt as belonging to the subspace of normal movements. Namely, the system statistically and adaptively learnt normal events in front of the camera just by using a lot of data. It should be noted that this is performed in unsupervised learning, viz. without any supervised learning about each movement.

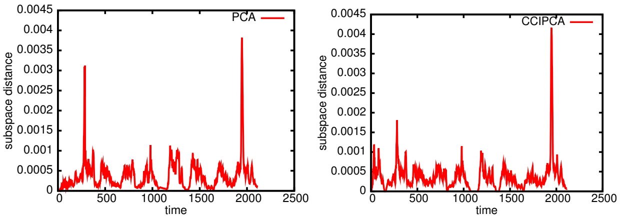
3.2 Online learning

Next, we experimented on online learning without dividing the phase. The data was the series of video sequences including multiple persons we used in the last subsection.

At first, we applied the method that solves the eigenvalue problem every step with updating auto-correlation matrix. **Figure 9 (a)** shows the abnormality value, the distance from the normal movement subspace constructed by using this method at each step. The abnormal value at the first several hundreds of frames were sometimes large and unstable, since the number of samples was too small to construct a statistically stable and valid subspace. As is seen, the value of an abnormal movement around at the 1950 frame was large, and an abnormal movement was successfully detected.

Secondly, we applied the other method using CCIPCA. **Figure 9 (b)** shows the distance from the subspace of normal movements constructed by using CCIPCA. Though this is the method to approximately construct the subspace of normal movements, an abnormal movement was successfully and more clearly detected.

Here, we set the rank of the subspace of normal movements four, since the rank was nearly constant through all the frames and the execution time for calculating the contribution rate was large. The execution time for estimating the contribution rate was 21 seconds per frame on a Pentium 4 3.02GHz, while by giving the rank the time became 0.0039 seconds per frame. For comparison, the time for solving eigenvalue problem at each step was 1.7 seconds per frame.



(a) PCA

(b) CCIPCA

Figure 9: Abnormality value. (online learning)

4 Conclusions

We showed that the present system can statistically detect abnormal movements even in scenes containing multiple persons. It should also be remarked that our method does not require the object segmentation nor tracking, and the computational amount is constant regardless of the number of persons.

Also, normal movements are not defined explicitly, whereas the system can learn normal movements and detect abnormal movements in quite an adaptive way such as the unsupervised learning of occlusions of persons. Further, we showed that the system can incrementally learn the normal movements and detect abnormal movements fast enough in real time.

In this paper, we experimented for the case that there was one type of normal movements in scenes. However, our method could be applied to multiple normal movements, provided that the abnormal movement feature vector is not represented as a linear combination of normal movement feature vectors (namely, linearly independent). If such a condition does not hold, we could improve the method by employing such as clustering in the subspace of normal movements. Thus, in any ways, we need to conduct more experiments for various scenes containing multiple types of normal movements.

In addition, our method is so general as not depending on objects, because no model of objects are assumed. Therefore, we need to evaluate its potential validity by performing experiments for other kinds of objects in various applications.

References

[1] A. A. Efros, A. C. Berg, G. Mori, and J. Malik, “Recognizing Action at a Distance,” *Proc. IEEE Int. Conf. on Computer Vision*, vol. 2, pp. 726-733, 2003.

[2] I. Haritaoglu, D. Harwood, and L. S. Davis, “W4: Who? When? Where? What? A Real Time System for Detecting and Tracking People,” *Proc. IEEE Int. Conf. on Face and Gesture Recognition*, pp. 222-227, 1998.

[3] W. Juyang, Z. Yilu, and WS. Hwang, “Candid Covariance-Free Incremental Principal Component Analysis,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 1034-1040, 2003.

[4] T. Kobayashi, and N. Otsu, “Action and Simultaneous Multiple-Person Identification Using Cubic Higher-Order Local Auto-Correlation,” *Proc. IAPR Int. Conf. on Pattern Recognition*, vol. 4, pp. 741-744, 2004.

[5] N. Otsu, and T. Kurita, “A New Scheme for Practical Flexible and Intelligent Vision Systems,” *Proc. IAPR Workshop on Computer Vision*, pp. 431-435, 1988.

[6] N. Otsu, “Discriminant and Least Squares Threshold Selection,” *4IJCPR*, pp.592–596, 1978.

[7] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, R. Williamson, “Estimating the Support of a High-Dimensional distribution,” *Neural Comput.*, vol. 13, no. 7, pp. 1443-1471, 2001.

[8] Y. Song, L. Goncalves, and P. Perona, “Unsupervised Learning of Human Motion,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 7, pp. 814-827, 2003.

[9] Y. Yacoob, and M. J. Black, “Parameterized Modeling and Recognition of Activities,” *Computer Vision and Image Understanding*, vol. 73, no. 2, pp. 232-247, 1999.

[10] L. Zelnik-Manor, and M. Irani, “Event-Based Analysis of Video,” *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. 123-130, 2001.

[11] H. Zhong, J. Shi, and M. Visontai, “Detecting Unusual Activity in Video,” *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. 819-826, 2004.