Incident Detection based on Dynamic Background Modeling and Statistical Learning using Spatio-temporal Features

Yasuhiro Murai, Hironobu Fujiyoshi Department of Computer Science, Chubu University. Matsumoto 1200, Kasugai, Aichi, Japan. yasu@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

Abstract

This paper presents a method for detecting an incident motion (e.g., tumble or violent action of a person) in a dynamic background scene. This method is based on the use of spatio-temporal features obtained using a space-time patch (ST-patch). Our approach consists of three steps: 1) dynamic background modeling using a Gaussian mixture model, 2) human regions detection based on Real AdaBoost, and 3) calculation of irregularity measure using weighted ST-patch features. The proposed method can be used to detect the incident motion in a scene with a dynamic background, which would be difficult to detect with a conventional method using cubic higher-order local auto-correlation (CHLAC) features. Our experimental results show that our method performs about 27% better than the conventional method in a contained scene with an escalator as the dynamic background.

1 Introduction

Video-understanding technologies, such as objects detection and tracking, have recently been used for large-scale surveillance camera systems. Today, incident detection (e.g., a person stumbling on an escalator) from a surveillance camera is becoming an important research topic for next-generation security systems.

As related works, Scholkopf *et al.* [1] proposed a method for abnormal motion detection that uses the probability distributions estimated using SVM, and Zhong *et al.* [2] proposed a method using the co-variance matrix of video stream features. Moreover, a method that uses CHLAC features has been proposed, which is one of the most typical methods used for abnormal motion detection.

The CHLAC features are extracted by applying a set CHLAC patterns to binary images obtained using temporal differencing. Thus, CHLAC features contain information on the "appearance" of "motion". Nanri *et al.* [3] proposed a method combining CHLAC features with a linear subspace method. This method achieved robust abnormal motion detection without using human detection and tracking. CHLAC features are also used for event detection [4].

However, these conventional methods have a problem. When a background is dynamic or there are objects that are not targets in the images, it is difficult to Masato Kazui Hitachi, Ltd., Hitachi Research Laboratory. 1-1, Omika 7, Hitachi, Ibaraki, Japan. masato.kazui.bq@hitachi.com

detect the incident motion. For example, the method that uses CHLAC features and the linear subspace method has difficulty in detecting motion because any incident motion that moves in the same direction as the dynamic background will be buried in the feature space according to the properties of the CHLAC with the additivity of the features.

To detect incident motion, we focus on the following two points: 1) removal of the dynamic background from an image, and 2) calculation of the irregularity measure of target object. To capture these two points, we propose a method that simultaneously uses spatiotemporal features containing information on both the "appearance" and "motion" for detecting irregular motions. This makes incident detection from a scene with a dynamic background scene possible.

This paper is structured as follows. Section 2 describes the properties of ST-patch which the proposed method uses, and Section 3 describes a procedure of incident detection using ST-patch features. We evaluate and compare our method with related work using an escalator scene in Section 4, and conclude in Section 5.

2 ST-patch

ST-patch features calculated from a ST-patch, which is a local region of images that extends in the time direction, are the spatio-temporal features containing information on both the "appearance" and "motion" simultaneously [5]. In addition, ST-patch features can calculate the irregularity measure of motion vectors. These ST-patch features are also used for event detection [6] for combined object detection and segmentation [7].

2.1 ST-patch Features

ST-patch features are extracted from the x, y, and t axis gradients in the local region of images that extends in the time direction (e.g., 25×25 [pixels]×5[frames]). Therefore, each pixel i has space-time gradients that contain three specific gradients, P_{x_i} , P_{y_i} , and P_{t_i} . By stacking all the n pixels within the ST-patch, we obtain

$$\mathbf{G} = \begin{bmatrix} P_{x_1} & P_{y_1} & P_{t_1} \\ P_{x_2} & P_{y_2} & P_{t_2} \\ \vdots & \vdots & \vdots \\ P_{x_n} & P_{y_n} & P_{t_n} \end{bmatrix}_{n \times 3}, \quad (1)$$



Figure 1: Procedure for proposed method.

where **G** denotes an $n \times 3$ matrix. By multiplying both sides of Eq.(1) by \mathbf{G}^{T} (the transpose of **G**), we can obtain $\mathbf{G}^{\mathrm{T}}\mathbf{G}$. We denote $\mathbf{G}^{\mathrm{T}}\mathbf{G}$ by **M**:

$$\mathbf{M} = \mathbf{G}^{\mathrm{T}}\mathbf{G} = \begin{bmatrix} \sum P_x^2 & \sum P_x P_y & \sum P_x P_t \\ \sum P_y P_x & \sum P_y^2 & \sum P_y P_t \\ \sum P_t P_x & \sum P_t P_y & \sum P_t^2 \end{bmatrix}.$$
 (2)

Matrix **M** is ST-patch features. This matrix contains information on both the "appearance" and "motion" simultaneously.

2.2 Continues Rank-Increase Measure

The continues rank-increase measure (CRIM) Δr is calculated from the eigenvalues of matrix **M** (see Eq. (2)) and matrix \mathbf{M}^{\diamond} :

$$\mathbf{M}^{\diamondsuit} = \begin{bmatrix} \sum P_x^2 & \sum P_x P_y \\ \sum P_y P_x & \sum P_y^2 \end{bmatrix}.$$
(3)

Let $\lambda_1 \geq \lambda_2 \geq \lambda_3$ be the eigenvalues of a 3×3 matrix **M**. Let $\lambda_1^{\diamond} \geq \lambda_2^{\diamond}$ be the eigenvalues of a 2×2 matrix \mathbf{M}^{\diamond} . From the interlacing property of the eigenvalues in the symmetric matrices it follows that $\lambda_1 \geq \lambda_1^{\diamond} \geq \lambda_2 \geq \lambda_2^{\diamond} \geq \lambda_3$. This leads to the following equation :

$$\Delta r = \frac{\lambda_2 \cdot \lambda_3}{\lambda_1^{\diamondsuit} \cdot \lambda_2^{\diamondsuit}}. \qquad (0 \le \Delta r \le 1) \tag{4}$$

 $\Delta r \approx 0$ is a consistent motion, and $\Delta r \approx 1$ is an inconsistent motion. Therefore we can calculate the measure of the irregular motion vectors within a ST-patch.

3 Incident Detection using ST-patch Features

To detect any incident motion, it is necessary to observe and recognize human motions. However, it is difficult to extract the motion features of humans if the background is dynamic or there are objects that are not the target objects in the images.

To make incident detection possible, we must construct models of the spatio-temporal features on dynamic backgrounds. Once the dynamic background is removed, and only the regions of an object can be detected. We then detect the human regions from the detected object regions using Real AdaBoost. Finally, we calculate the irregularity measure of the motion vectors obtained from the detected human regions. In this way, incident detection is possible from a scene with a dynamic background. The procedures for the proposed method are illustrated in Fig.1.

3.1 Dynamic Background Modeling using Gaussian Mixture Model

We use a Gaussian mixture model (GMM) for modeling the observed ST-patch features from a dynamic background. The GMM is a probability model that combines multiple Gaussian distributions [8]. Let $p(\boldsymbol{x})$ be the probability density, where \boldsymbol{x} is data, and let $p(\boldsymbol{x}|i)_{i=1,...,C}$ be the probability density with Cclasses. Let $p(\boldsymbol{x})$ be the weighted linear combination of $p(\boldsymbol{x}|i)_{i=1,...,C}$, which is expressed by

$$p(\boldsymbol{x}) = \sum_{i=1}^{C} w_i p(\boldsymbol{x}|i), \qquad (5)$$

where

$$p(\boldsymbol{x}|i) = \frac{1}{\sqrt{(2\pi)^{M}|\boldsymbol{\sigma}_{i}|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{i})^{\mathrm{T}}\boldsymbol{\sigma}_{i}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_{i})\right\}, \quad (6)$$

and M is the number of dimensions, w_i is the weight, μ_i is the average, and σ_i is the covariance matrix. In this way, GMM needs three parameters $(w_i, \mu_i,$ and $\sigma_i)$ for each Gaussian distribution. Thus we estimate each parameter using a deterministic annealing EM algorithm[9]. In addition, the proposed method updates the online of the Gaussian mixture models by applying an adaptive mixture of Gaussian background models[10].

ST-patch features have the same values in the symmetrical component. Thus, the proposed method is represented as a 6-dimensional vector \mathbf{v} as follows :

$$\mathbf{v} = \left(\sum P_x^2, \sum P_x P_y, \cdots, \sum P_t^2\right).$$
(7)

The proposed method can be used to create dynamic background models by applying GMM to ST-patch features denoted as a 6-dimensional vector **v**. We distinguish an object/background region by comparing the Gaussian mixture model with input features. By comparing the Gaussian mixture model to an ST-patch arranged in a grid pattern (e.g., the ST-patch size is set to 25×25 [pixels] $\times 5$ [frames], shift of the ST-patch is set to 5[pixels].), we can remove the dynamic background.

3.2 Human Region Detection based on Real AdaBoost

The dynamic background modeling method cannot be used to specify human regions. Recently, many methods for combining low-level features with a statistical learning method have been proposed as an approach for the object detection in an image. One of the successful methods is a people detection algorithm



Figure 2: Example of learning samples.

using a dense grid of histograms of oriented gradients (HOG) [11], and a face detection system using appearance patterns obtained using Haar-like features [12]. In this paper, we propose a method for human region detection that combines ST-patch features and a statistical learning method.

We use Real AdaBoost [13] as the statistical learning method. Real AdaBoost is a method of calculating the degree of similarity from the probability density distribution of each dimension in the features of a positive class and negative classes. Then, it learns the features most effective in discernment. We uses a human region of the normal motion as the positive class and other region as negative class. Example of learning samples are illustrated in Fig.2. The proposed method learned STpatch features calculated from these learning samples using Real AdaBoost. An overview of the construction of a human region classifier is illustrated in Fig.3.

ST-patch When size is set to 25X 25[pixels]×5[frames], the shift of the ST-patch is set to 5[pixels], and a detection window is set to 140×260 [pixels] (the number of ST-patches that exists in the window is 1,081.), the spatio-temporal features from the region of a detection window have a $1,081\times 6=6,486$ dimensional vector. The proposed method learns these features using Real AdaBoost and constructs a classifier. When we use Real AdaBoost, it learns the features obtained from the images in which removed the dynamic background.

If we use the ST-patch features for human region detection, a detection rate will decrease by affecting clothes or lighting change. Therefore, we normalize ST-patch features \mathbf{v} (see Eq. (7)), expressed as a 6-dimensional vector using all the vectors. We denote the normalized ST-patch features by $\mathbf{v}_{\rm N}$:

$$\mathbf{v}_{\mathrm{N}} = \frac{v}{\left(\sum_{i=1}^{6} |v_i| + \epsilon\right)},\tag{8}$$

where ϵ is a small positive constant. The proposed method uses the normalized feature vectors $\mathbf{v}_{\rm N}$. By learning the normalized feature vectors, we can construct a classifier, which is not affected by clothes or lighting change.

To detect human regions using the constructed classifier, we conduct a raster scanning of the detection windows over the image from which the dynamic background was removed. We assume that the detection accuracy will be improved using these images. We try



Figure 3: Construction overview of human region classifier.

to determine whether the features inside the detection window are human regions or not. Finally the windows are unified using mean-shift clustering[14].

3.3 Incident Detection by Irregularity Measure of Motion Vectors

In general, incident motion, such as a person stumbling, makes the motion vectors change a lot compared to normal motion. Therefore, we use the CRIM (see 2.2) calculated from ST-patch features as the judgment index for determining normal or incident motion.

3.3.1 CRIM using Weighted ST-patch Features

Real AdaBoost expresses the probability density distribution by computing a histogram from the features, and constitutes the look-up table (LUT) weak classifier. Therefore, we can obtain the real values which express the confidence values of input features from the output of the LUT weak classifier. The proposed method computes the CRIM using ST-patch features weighed by the output value of the LUT weak classifier. Thereby we assure improvement in incident detection performance.

3.3.2 Calculation of Weighted ST-patch Features

We calculate the CRIM using the weighted space-time gradients obtained from Real AdaBoost. When detecting a human region based on Real AdaBoost, we obtain a weak classifier output

$$h(x) = \frac{1}{2} \ln \frac{W_{+}^{j} + \epsilon}{W_{-}^{j} + \epsilon}, \qquad (9)$$

and a separation distance of probability density distribution

$$\alpha = 1 - \sum_{j} \sqrt{W_{+}^{j} W_{-}^{j}}, \qquad (10)$$

where j is the number of bins of a 1-dimensional histogram corresponding to the features, W^j_+ and W^j_-



Figure 4: Evaluated value.

are the probability density of the positive and negative classes, and ϵ is a small positive constant. We can obtain the real value from the weak classifier output.

We calculate the evaluated values by multiplying h(x) by α . The proposed method uses the evaluated values only when h(x) > 0. Therefore, we denote the evaluated value by E:

$$E = \begin{cases} h(x)\alpha & \text{if } h(x) > 0\\ 0 & \text{otherwise} \end{cases}$$
(11)

Thus, we can obtain the distribution illustrated in Fig.4 by calculating the evaluated values. By using these values with this information, we can obtain weighted space-time gradients, which suppress the effect of the dynamic background and accentuate the motion vectors within the local regions of an object (Fig.5).

We calculate the matrix of the weighted space-time gradients by multiplying the matrix \mathbf{M} (see Eq. (2)) by E (see Eq. (11)). We express this as \mathbf{G}_E :

$$\mathbf{G}_{E} = \begin{bmatrix} E_{1}P_{x_{1}} & E_{1}P_{y_{1}} & E_{1}P_{t_{1}} \\ E_{2}P_{x_{2}} & E_{2}P_{y_{2}} & E_{2}P_{t_{2}} \\ \vdots & \vdots & \vdots \\ E_{n}P_{x_{n}} & E_{n}P_{y_{n}} & E_{n}P_{t_{n}} \end{bmatrix}_{n \times 3} . (12)$$

From matrix \mathbf{G}_E , which is weighted to the space-time gradients using the evaluated values, we calculate STpatch features with the same approach as 2.1, and we calculate the CRIM using the same approach as 2.2.

4 Evaluation Experiment

We experimented with incident detection to evaluate the validity of the proposed method. To compare the proposed method with the conventional method, we used escalator surveillance images containing a dynamic background. We used 6430 positive samples and 2844 negative samples.



Figure 5: Weighted space-time gradients.

We evaluate the following two approach for the proposed methods.

- Proposed method1 : a method using the CRIM calculated from the weighted ST-patch features (Weighted ST-patch features).
- Proposed method2 : a method using the CRIM calculated from the non-weighted ST-patch features (ST-patch features).

We also evaluate the following two approach for the conventional methods.

- Conventional method1 : a method using CHLAC features[3] (CHLAC features).
- Conventional method2 : a method using CHLAC features calculated from images which the dynamic background has been removed (Removal of dynamic background + CHLAC features).

4.1 Experimental Results

Fig.6 shows examples of incident detection using the proposed method. As shown in Fig.6, the proposed method can be used to detect the incident motion from a scene with a dynamic background containing several people. This is because the proposed method detects human regions using ST-patch features and calculates the irregularity measures for each human region.

4.1.1 Results of Proposed Method (ST-patch Features)

Incident detection results using the proposed method1 and the proposed method2 are illustrated in Fig.7. As shown in Fig.7, we can see that the proposed method1 distinguishes between normal and incident motion when the threshold of the CRIM is set to TH=0.7. The proposed method1, compared with the proposed method2, produces a higher CRIM during incident motions. And the proposed method1 is observing a lower CRIM during normal motions. Therefore, we verified that the proposed method1 performances better than the proposed method2.

In proposed method2, the CRIM is calculated from the whole human region including the dynamic background. As a result, CRIM becomes a global value that contains the dynamic background information. On the other hand, in the proposed method1, the CRIM can be used to suppress the effect of dynamic background and accentuating the motion vectors within the local regions of an object. Therefore, it was able to improve detection performance.



Figure 6: Example of incident detection using proposed method.

Proposed method1 : Weighted ST-patch features



Figure 7: Incident detection results using ST-patch features.

4.1.2 Results of Conventional Method (CHLAC Features)

Incident detection results using the conventional method1 and the conventional method2 are illustrated in Fig.8. As shown in Fig.8, we can see that the conventional method1 seldom distinguishes between normal and incident motion. On the other hand, the conventional method2, compared with the conventional method1, distinguishes between normal and incident motion because it does not affect the dynamic background. However, it may falsely detect by crossing the person.

4.2 Quantitative Evaluation

To quantitatively evaluate the proposed method, we used a receiver operating characteristic (ROC) curve. The evaluation data sets that were used were 46 scenes chosen from escalator surveillance images. They contain 28 scenes of normal motion and 18 scenes of abnormal motion.

Fig.9 shows the ROC curves of each method. As shown in Fig.9, the proposed method1 is more efficient than the other methods. Moreover, the proposed method1 is about 27% more efficient than the conventional method1. Therefore, we can obtain the high performance by suppressing the effect of the dynamic background and accentuating the motion vectors within the local regions of an object.

5 Conclusion

In this paper, we proposed a method for removing the dynamic background, detecting the human regions based on Real AdaBoost, and also detecting the incident motion. The proposed method could detect the incident motion, even in a scene that had an escalator as the dynamic background. Our experimental results proved the validity of our proposed method.

Our future works will include improving the accuracy of detecting human regions and supporting multiscale human region detection. In the proposed method, the accuracy of human region detection is important because the incident motion is distinguished for every detected human region. However, in a scene with an overlap of people, detecting each human region becomes difficult. Therefore, improvement in human region detection accuracy is required. The proposed method does not support a multi-scale human region because it fixes the size of a detection window. In our future work, we will attempt to detect a multi-scale human region using adaptive transformation of the ST-



Figure 8: Incident detection results using CHLAC features.



Figure 9: Quantitative evaluation using ROC curves.

patch size.

References

- B. Scholkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson. Estimating the support of a highdimensional distribution. *Neural Comput*, Vol. 13, No. 7, pp. 1443–1471, 2001.
- [2] H. Zhong, J. Shi, and M. Visontai. Detecting unusual activity in video. In Proc. Computer Vision and Pattern Recognition, Vol. 2, pp. 819–826, 2004.
 [3] T. Nanri and N. Otsu. Anomaly detection in motion
- [3] T. Nanri and N. Otsu. Anomaly detection in motion images containing multiple persons. *IPSJ Transactions on Computer Vision and Image Media*, Vol. 45, No. SIG.15, pp. 43–50, 2005.

- [4] T. Kobayashi and N. Otsu. Action and simultaneous multiple-person identification using cubic higher-order local auto-correlation. In Proc. International Conference on Pattern Recognition, pp. 741–744, 2004.
- [5] E. Shechtman and M. Irani. Space-time behavior based correlation -or- how to tell if two underlying motion fields are similar without computing them? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 11, pp. 2045–2056, 2007.
- [6] R. Sukthankar Y. Ke and M. Hebert. Event detection in crowded videos. In Proc. International Conference on Computer Vision, pp. 8–15, 2007.
- [7] Y. Murai, H. Fujiyoshi, and T. Kanade. Classifying direction of movement and segmenting regions of objects by using space-time patches. *IPSJ Transactions* on Computer Vision and Image Media, Vol. 1, No. 2, pp. 21–31, 2008.
- [8] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, Vol. 39, No. 1, pp. 1–38, 1977.
- [9] N. Ueda and R. Nakano. Deterministic annealing em algorithm. *IEICE Transactions on Information and Systems*, Vol. J80-D-II, No. 1, pp. 267–276, 1997.
- [10] A. Shimada, D. Arita, and R. Taniguchi. Increment and decrement of gaussians in adaptive mixture-ofgaussian background models. 9th Meeting on Image Recognition and Understanding, pp. 746–751, 2006.
- [11] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In Proc. Computer Vision and Pattern Recognition, Vol. 1, pp. 886–893, 2005.
- [12] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, Vol. 57, No. 2, pp. 137–154, 2004.
- [13] B. Wu, H. Ai, C. Huang, and S. Lao. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, No. 37, pp. 297–336, 1999.
- [14] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans*actions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 5, pp. 603–619, 2002.