Abnormal Action Detection Using Temporal and Spatial Information

Hisashi Mikado and Kazunori Onoguchi
Graduate School of Science and Technology, Hirosaki University, Hirosaki, Japan
onoguchi@eit.hirosaki-u.ac.jp

Abstract

This paper presents a method to detect abnormal action by using spatio-temporal features. At first, silhouette images obtained from the background subtraction are separated into the upper part and the lower part. Each part is divided into some blocks so that the number of pixels in each block may be distributed equally. Next, motion features are detected by applying the Fourier transform to time series data consisting of the pixel density in each block. In addition to motion features, shape features are detected by calculating the mean and the variance of the pixel density in each block. The classifier using one class SVM is created by learning feature vectors obtained from silhouettes created by 3D CG software. The proposed method can detect both dynamic abnormal action and static abnormal posture. In experiments, walking, running or standing was identified as normal action or normal posture and the other was identified as abnormal action. The validity of the proposed method was confirmed by experimental results.

1. Introduction

Recently, the expectation for automatic surveillance is growing as many surveillance cameras have been arranged in various places. A lot of methods were proposed for cost reduction, evasion of a human mistake, early detection of abnormal situation and so on. Especially, it's important to detect abnormal situation because of preventing a crime or an accident and various methods have been proposed according to the use.

Seki et.al.[1] proposed the method to detect anomalous pattern of elderly people by the distance between images. Sumi et.al.[2] detected the violence in an elevator cage from the variance of the optical flow. Murai et.al.[3] detected abnormal behavior around the escalator by using Space Time Patch. These methods are developed only for the specific use, respectively. Ivanchenko et.al.[4] proposed the real-time method to detect abnormal motion by using macroblock motion vectors obtained from MPEG images. Bauckhage et.al.[5] detected abnormal gait from the frequency feature. These method cannot detect abnormal situation whose motion is small. Nanri et.al.[6] proposed abnormal detection method using CHLAC features[7]. This method can identify an abnormal action without extracting person's area and tracking it individually. However, this method cannot specify where abnormal action occurred in an image because it uses CHLAC features created from an entire image. Moreover, computational complexity of this method is great and parallel processing is required for real-time processing. Boiman et.al.[8] detected irregularity by comparing the space-time patch obtained from intensity gradient with the database. Although this method can detect both dynamic abnormalities and static abnormalities, it needs huge data base.

This paper proposes the method which can detect both dynamic abnormal action and static abnormal posture although computational cost is small and the size of the data base used for identification is small. Our method can also specify where the abnormal behavior occurs in an image. The frequency feature for a motion and the shape feature for a posture are extracted from the pixel density of a person's silhouette in time series images. Our method is robust to individual difference of action and posture because the pixel density absorbs it. The proposed method assumes that human's normal action is walking, running or standing. Except these actions, it is judged to be abnormal. Each action is identified by one class SVM learned by silhouettes created by 3D CG software.

2. Detection of Abnormal action

The proposed method discriminates a person's action from a silhouette because the change of a silhouette contains much information of actions. At first, a person's silhouette is divided into the upper part and the lower part. Each part is divided into some blocks so that the number of silhouette pixels in each block may be distributed equally. Next, the motion feature is detected by applying the Fourier transform to time series data consisting of the pixel density in each block. The shape feature is detected by calculating the mean and the variance of the pixel density in each block. The classifier using one class SVM is created by learning feature vectors consisting of motion features and shape features. Each action is identified by this classifier.

2.1 Silhouette tracking

A person's silhouette is extracted by the background subtraction. Because the background subtraction for each pixel is weak to the image noise, average intensity of pixels around each point is used for subtraction. Each silhouette is tracked between frames by the mean shift based on the color histogram.

2.2 Pixel density of silhouette

At first, a person's silhouette is divided into the upper part and the lower part so that the number of pixels in each part may become the same.

$$UP = LP = \frac{P}{2} \tag{1}$$

In eq. 1, P is the total number of pixels contained in a person's silhouette, UP is the number of pixels contained in the upper part and LP is the number of pixels contained in the lower part.

Next, a rectangle which circumscribes a silhouette is obtained in each part. The lower part may contain a part of upper half of body, such as hands, clothes and so on, as shown in Fig. 1 (a). Therefore, a rectangular left end and right end of the lower part is estimated from the lower region as shown in Fig. 1 (b) and (c). The upper part and the lower part are divided into n blocks respectively so that the number of pixels of a silhouette may be distributed equally in each block as shown in eq. 2.

$$sUP = \frac{UP'}{n} \quad (sLP = \frac{LP'}{n}), \tag{2}$$

where sUP is the number of pixels in each block of the upper part, sLP is the number of pixels in each block of the lower part, UP' is the number of pixels in the upper part, LP' is the number of pixels in the lower part and n is the number of blocks in each part. Figure 2 shows some examples of blocks divided based on the eq. 2.

In each block, the pixel density is calculated by

$$UD_i = \frac{sUP}{US_i} \quad (LD_i = \frac{sLP}{LS_i}), i = 1, \cdots, n, \tag{3}$$

where UD_i is the pixel density in block i of the upper part, LD_i is the pixel density in block i of the lower part, US_i is the area of block i of the upper part and LS_i is the area of block i of the lower part.

(a) (b) (c)

Figure 1. Rectangle of lower part

2.3 Motion feature

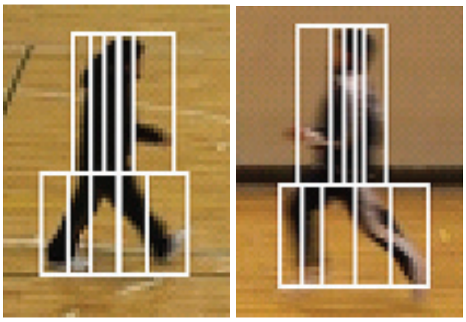
The time series data consisting of pixel density of the T frames are created in each block. The motion feature is detected by applying FFT (Fast Fourier Transform) to this time series data. A frequency component is considered to be a level from 1 to $\frac{T}{2}$ in ascending order of frequency. The frequency level with the maximum energy from the power spectrum is selected as the motion feature in each block. In the example of Fig. 3, the level 2 is selected as the motion feature. If the energy of the selected level is lower than the predetermined threshold α, its level is set as 0 because this block contains no motion area. The motion feature consists of a frequency level obtained in each block. Therefore, the dimension of the motion feature is $2n$. Figure 4 shows examples of motion features. In each block, higher frequency level is expressed as the brighter color. The action with no periodic motion shown in Fig. 4 (c) shows various frequency levels in each block. On the other hand, periodic action such as walking or running shows a single frequency level as shown in Fig. 4 (a) and (b).

2.4 Shape feature

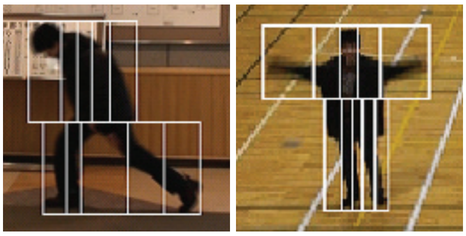
In order to detect static abnormal action, the following three shape features are calculated in each frame.

1. The pixel density of each block ($2n$ dimensions)
2. The normalized height and width of the rectangular lower part and upper part (4 dimensions)
3. The aspect ratio of a rectangle which circumscribes a person's silhouette (1 dimension)

The mean and variance of each feature for T frames is calculated as the shape feature of a person's silhouette. The combined vector of the motion feature and the shape feature is utilized for the abnormal detection. The total dimension of the feature vector is $2n + (2n + 4 + 1) \times 2 = 6n + 10$.



(a) Walk (Normal) (b) Run (Normal)



(c) Drag the leg (Abnormal) (d) Swing hands (Abnormal)

Figure 2. Blocks on a person's silhouette

2.5 Learning and discrimination

The classifier using one class SVM is created by learning the feature vectors. RBF kernel is used for learning. The classifier identifies whether an action is normal or abnormal for every frame after the T frame. In order to avoid the incorrect identification which occurs temporarily by an image noise and so on, the identification result for 60 frames is accumulated and it is judged that a person's action is abnormal when the frame identified as it's abnormal is more than 40% or more. When people overlap in an image, it may be identified as abnormality because the shape of person's silhouette becomes less normal. To avoid this situation, the threshold for identification is raised to more than 80% when persons' silhouettes overlap each other. The threshold is returned to 40% after persons' silhouettes separate and 60 frames passes.

In the learning phase, the proposed method uses silhouette images of person's action created by 3D CG software (POSER). Silhouette images for walking, running and standing are created while changing the person's moving direction and the tilt of a camera. The person's moving direction is from 0 to 360 degrees and the tilt of a camera is 0, 15, 30 or 45 degrees from a horizontal direction to the downward.

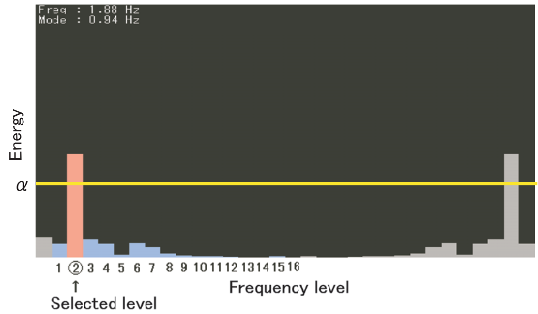
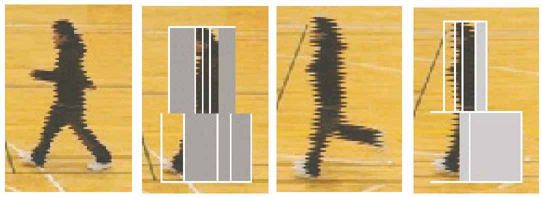
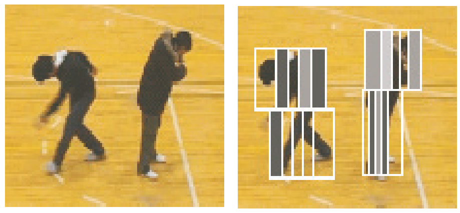


Figure 3. Power spectrum



(a) Walk (Normal) (b) Run (Normal)



(c) Fight (Abnormal)

Figure 4. Frequency level in blocks

3 Experiments

We conducted experiments which detect abnormal action in indoor and outdoor scenes. An image size is 320×240 pixels and the processing time is $30fps$ on PC with Core2Extreme 3.0GHz CPU. 25 abnormal actions and 20 normal actions are used in the test phase. The dimension of the feature vector is 40 because the number of blocks n is 5. The number of frames T is 32.

Figure 5 shows fighting persons and Fig. 6 shows a stumbling person. Persons identified as abnormal action are surrounded by the red ellipse. The proposed method was able to distinguish abnormal actions and normal actions even if several persons exist in an image. Figure 7 shows a person who are dragging a leg. Although the distance and the viewing direction to a person is different from these of Figs.5 and 6, this scene was identified as the abnormal action correctly because the pixel density is not subject to the influence of size and various silhouettes created by 3D CG are used for learning. In Fig.8, the person who crouched down was detected as the abnormal action.

20 normal actions were identified correctly. However, 4 in 25 abnormal actions were wrongly detected as a normal action. These were abnormal actions with small motion and similar shape to standing posture, such as looking around restlessly.

4 Conclusion

This paper proposed the method which can detect both dynamic abnormal action and static abnormal posture. The proposed method identifies abnormal action and posture at a video rate because the simple feature based on the pixel density is used for identification. Experimental results show the effectiveness of the proposed method. It is a future work to improve detection accuracy.

References

[1] H. Seki and Y. Hori, "Detection of Abnormal Human Action Using Image Sequence," Proceedings of International Power Electronics Conference, pp.1272-1277, 2000.

[2] K. Sumi, M. Seki and H. Shiozaki, "Video Alalysis Technology for Elevator Cage Abnormality Detection," IPSJ Magazine, Vol.48, No.1, pp.17-22, 2007 [in Japanese].

[3] Y. Murai, H. Fujiyoshi and M. Kazui, "Incident Detection based on Dynamic Background Modeling and Statistical Learning using Spatio-temporal Features," Proceedings of MVA2009,pp.156-161,2009.

[4] Y. Ivanchenko, S. Rochel, N. Kiryati, T. R. Raviv, "Real-time abnormal motion detection in surveillance video," Procedings of ICPR2008, pp.1-4, 2008

[5] C. Bauckhage, J.K. Tsotsos, F.E. Bunn, "Automatic detection of abnormal gait," Image and Vision Computing, Vol.27, pp. 108-115, 2009.

[6] T. Nanri and N. Otsu, "Abnormaly Detection in MOtion Images Containing Multiple persons," IPSJ CVIM46 (SIG15(CVIMpp.43-50, 2005 [in Japanese].

[7] N. Otsu, "Towards Flexible and Intelligent Vision Systems - From Thresholding to CHLAC-", Proceedings of MVA2005, pp430-439, 2005.

[8] O. Boiman and M. Irani, "Detecting irregularities in images and in video," Proceedings of ICCV2005, 2005.

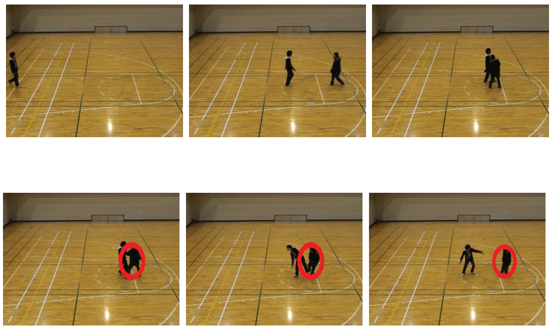


Figure 5. Experimental result (Fight)

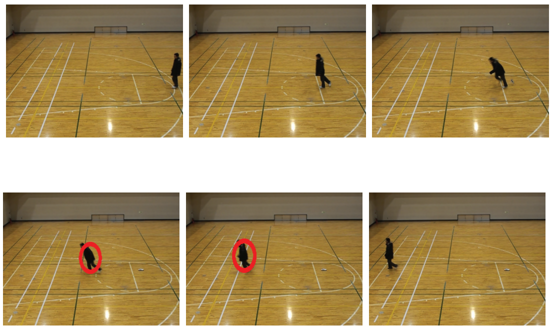


Figure 6. Experimental result (Stumble)

Figure 7. Experimental result (Dragging a leg)

Figure 8. Experimental result (Crouching down)