2013 Fourth International Conference on Emerging Security Technologies

Abnormal behavior detection using privacy protected videos

Yumi Iwashita, Shuhei Takaki, Ken'ichi Morooka, Tokuo Tsuji, Ryo Kurazume
Information Science and Electrical Engineering, Kyushu University
Fukuoka, Japan 819-0013
Email: yumi@ieee.org, takaki@irvs.ait.kyushu-u.ac.jp,
{morooka, tsuji, kurazume}@ait.kyushu-u.ac.jp

***Abstract*—Intelligent visual surveillance, which relies heavily on human motion detection / recognition and people recognition, has received a lot of attention for its use in effective monitoring of public places. However, there is a concern of loss of privacy due to distinguishable facial and body information. To deal with this issue, researchers proposed to protect privacy example by filtering of face or body areas, and developed methods of people identification from videos in which people's faces has been obfuscated, masked by digital filters. Along the same line of research dealing with videos in which the people faces were masked by filters, this paper introduces a method to detect abnormal behavior.**

In the proposed method, we first mask face areas in videos by Multiple Instance Learning tracking, and extract silhouette area from each image. We then extract features using affine moment invariants, and perform classification. We build a database including normal and abnormal behaviors, and we show the effectiveness of the proposed method on cases from the database.

I. Introduction

Ensuring security with intelligent visual surveillance, which relies heavily on human motion detection / recognition and people recognition, has proven very effective in monitoring public places, such as stations, airports, shopping malls, etc., as well as for facilitating healthcare services, e.g. in the monitoring of activities of old people [1]. Although video surveillance's usefulness in repressing crime is extremely valuable there is an concern of loss of privacy due to distinguishable facial and body information [2]. Research have addressed this issue by using for example, filters to block/mask the face or specific body areas [3] [4]. People identification in videos with masked face has been proposed in [5].

This paper, introduces a method to detect abnormal behavior using videos in which people's faces were masked. The target application is usage in a hospital or a healthcare facility. In the proposed method, at first we detect face areas in videos and extract silhouette area from each image. Next, extract features using affine moment invariants, which are used in person identification [6], object recognition [7], followed by a classification. We build a database including normal and abnormal behaviors, and we show experiments performed with the database.

II. Behavior classification

In this section firstly we explain the way to detect face areas in videos and to extract silhouette area from each image. Next, we explain a method to extract features using affine moment invariants, followed by an abnormal behavior detection.

A. Face detection and silhouette extraction

Figure 1 (a) shows example images of an abnormal behavior sequence which we utilize in experiments (details in Section 3), and it is clear that face area of a person is often occluded and its appearance changes a lot due to varieties of motions. To track his face stably even with his various motions, we utilize MILTrack [8], which can track an object in a video given its location in the first frame and no other information. This method updates an adaptive appearance model of a tracking system, using Multiple Instance Learning (MIL) to train the appearance classifier and an online boosting algorithm for MIL. In the MILTrack, a face location in the first frame is detected with Haar-like features, but in videos, which we utilize in experiments, it is often hard to detect the face location due to occlusion / limited face area. Thus in our system we manually detect the face location in the first frame. Figure 1 (b) shows example results of detected face area.

A background subtraction method is applied to captured images with masked face. Figure 1 (c) shows examples of silhouette areas.

B. Feature extraction

Affine moment invariants are extracted from each silhouette image as features [6]. In [6], we utilized for gait-based person identification, and we showed that affine moment invariants had almost the same discrimination capability with an gait energy image (GEI) [9] and an active energy image (AEI) [10] even though the number of features of affine moment invariants is much smaller than those [9] [10].

Affine moment invariants are moment-based descriptors, which are invariant under a general affine transform. The moments describe shape properties of an object as it appears. For an image the centralized moment of order $(p+q)$ of an object O is given by

$$\mu_{pq} = \sum \sum_{(x,y)\in O} (x-x_g)^p (y-y_g)^q I^{average}(x,y). \quad (1)$$

Here, x_g and y_g are the center of the object. The number of affine moment invariants $\boldsymbol{A} = \{A_1, A_2, \ldots, A_M\}$ is M, and

978-0-7695-5077-0/13 $26.00 © 2013 IEEE
DOI 10.1109/EST.2013.16

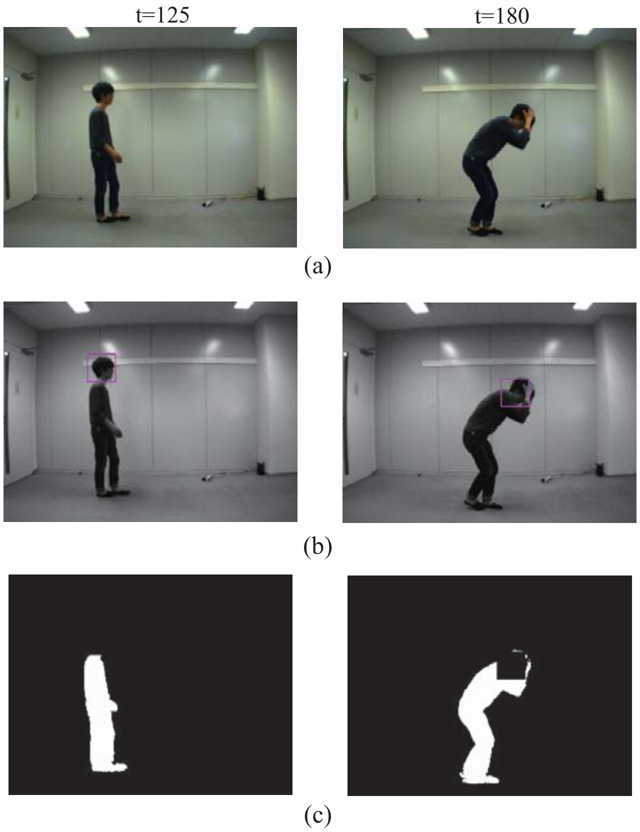

Fig. 1. (a) Examples of captured images, (b) detected face area (with pink rectangular), (c) silhouette images

we show six of them [11].

$$
\begin{aligned}
A_1 &= \frac{1}{\mu_{00}^4}(\mu_{20}\mu_{02} - \mu_{11}^2) \\
A_2 &= \frac{1}{\mu_{00}^{10}}(\mu_{30}^2\mu_{03}^2 - 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} \\
&\quad +4\mu_{30}\mu_{12}^3 + 4\mu_{03}\mu_{21}^3 - 3\mu_{21}^2\mu_{12}^2) \\
A_3 &= \frac{1}{\mu_{00}^7}(\mu_{20}(\mu_{21}\mu_{03} - \mu_{12}^2) - \mu_{11}(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}) \\
&\quad +\mu_{02}(\mu_{30}\mu_{12} - \mu_{21}^2)) \\
A_4 &= \frac{1}{\mu_{00}^{11}}(\mu_{20}^3\mu_{03}^2 - 6\mu_{20}^2\mu_{11}\mu_{12}\mu_{03} - 6\mu_{20}^2\mu_{02}\mu_{21}\mu_{03} \\
&\quad +9\mu_{20}^2\mu_{02}\mu_{12}^2 + 12\mu_{20}\mu_{11}^2\mu_{21}\mu_{03} \\
&\quad +6\mu_{20}\mu_{11}\mu_{02}\mu_{30}\mu_{03} - 18\mu_{20}\mu_{11}\mu_{02}\mu_{21}\mu_{12} \\
&\quad -8\mu_{11}^3\mu_{30}\mu_{03} - 6\mu_{20}\mu_{02}^2\mu_{30}\mu_{12} + 9\mu_{20}\mu_{02}^2\mu_{21}^2 \\
&\quad +12\mu_{11}^2\mu_{02}\mu_{30}\mu_{12} - 6\mu_{11}\mu_{02}^2\mu_{30}\mu_{21} + \mu_{02}^3\mu_{30}^2) \\
A_5 &= \frac{1}{\mu_{00}^6}(\mu_{40}\mu_{04} - 4\mu_{31}\mu_{13} + 3\mu_{22}^2) \\
A_6 &= \frac{1}{\mu_{00}^9}(\mu_{40}\mu_{04}\mu_{22} + 2\mu_{31}\mu_{22}\mu_{13} - \mu_{40}\mu_{13}^2 \\
&\quad -\mu_{04}\mu_{31}^2 - \mu_{22}^3)
\end{aligned}
\tag{2}
$$

C. *Abnormal behavior detection*

We use the nearest neighbor as the classifier. In the training phase, we extract features by the affine moment invariants from silhouette images, and then build a database. Then in the detection phase, features are extracted from silhouette images of a subject and the subject's abnormal motion is detected by the classifier.

III. Experiments

This section presents the results of behavior detection experiments performed using a database consisting of two kinds of behaviors, normal walking and abnormal behavior. We intend to use the proposed system in a hospital, and the abnormal behavior in the database includes (i) sitting down, (ii) falling, (iii) bringing one's hands to his chest and (iv) grabbing his head with his hands. Figure 2 shows examples of each behavior. From these examples, it is clear that a subject's motion varies and his appearance changes a lot due to varieties of motions. The image resolution is 775×600 and the number of subjects is 5, 4 sequences for each motion. Totally there are 100 sequences ($=5\times5\times4$) in the database, and the total number of images is around 8000.

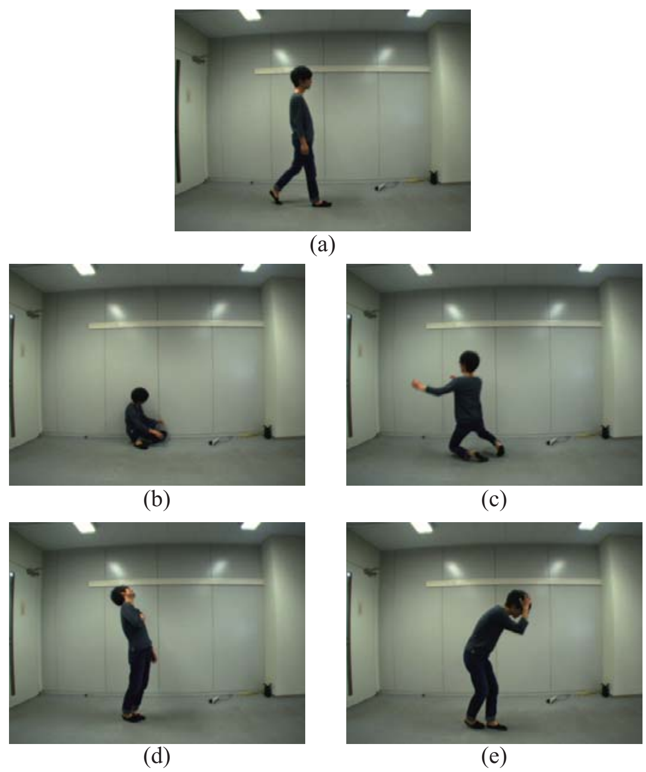

Fig. 2. (a) Normal walking, (b) sitting down, (c) falling, (d) bringing one's hands to his chest and (e) grabbing his head with his hands

We applied the proposed method to the database for evaluation. Here, we eliminated all motions of a subject, which are used as test datasets, from training datasets. In case that a motion is categorized as one of the abnormal motions, we classified the motion as an abnormal behavior.

TABLE III. CONFUSION MATRIX FOR EACH MOTION IN ABNORMAL BEHAVIOR (IMAGE RESOLUTION IS 775×600) [%]

	(i) Sitting down	(ii) Falling	(iii) Bringing one's hands to his chest	(iv) Grabbing his head with his hands
(i) Sitting down	60 %	15 %	20 %	0 %
(ii) Falling	15 %	30 %	30 %	15 %
(iii) Bringing one's hands to his chest	10 %	15 %	45 %	25 %
(iv) Grabbing his head with his hands	20 %	5 %	25 %	50 %

TABLE I. CORRECT CLASSIFICATION RATES OF (A) NORMAL WALKING ONLY, (B) ABNORMAL BEHAVIOR ONLY [%]

Image resolution	775×600 (100%)	387×300 (50%)	77×60 (10%)
(a) normal walking	90 %	90 %	85 %
(b) abnormal behavior	97.5 %	100 %	100 %

TABLE II. CORRECT CLASSIFICATION RATES OF EACH ABNORMAL BEHAVIOR [%])

Image resolution	775×600 (100%)	387×300 (50%)	77×60 (10%)
(i) Sitting down	60 %	60 %	70 %
(ii) Falling	30 %	30 %	20 %
(iii) Bringing one's hands to his chest	45 %	35 %	45 %
(iv) Grabbing his head with his hands	50 %	60 %	65 %

Table I shows the correct classification rates of normal walking and abnormal behavior, respectively. This also shows the results with different image resolution, since image resolution can change due to the distance between the camera and subjects. From these results, the correct classification rates of normal walking get low with the decrease of image resolution, and it tends to be categorized as abnormal behavior. On the other hand, the correct classification rates of abnormal behavior get better. One of reasons can be follows. The distribution of features of abnormal walk may be much bigger than that of normal walk, and the features of low image resolution tend to be close to those of abnormal behavior.

After the detection of abnormal behavior, we classified each behavior, which categorizes as abnormal behavior, as one of four behaviors. Table II shows the correct classification rate of each abnormal behavior, and Table III shows confusion matrix for abnormal behavior. From these results, it is clear that each abnormal motion is often misclassified with different abnormal motion.

IV. CONCLUSION

This paper presented a method to detect abnormal behavior using videos with masked face. Experimental results showed that the performance was high with high resolution images, but there was a tendency that normal behavior was misclassified as an abnormal one in case of low resolution images. Future work will include to develop a method which utilizes spatio-temporal features, which we did not use in our method, to achieve higher discrimination capability, and compare the method with state-of-the-art methods.

ACKNOWLEDGMENT

This work is supported by Strategic Information and Communications R&D Promotion Program (SCOPE) No.121810005.

REFERENCES

[1] O. Popoola & K. Wang, *Video-Based Abnormal Human Behavior Recognition: A Review*, SMC-C(42), No. 6, pp.865-878, 2012.

[2] A. Cavallaro, *Privacy in Video Surveillance*, IEEE Signal Proc. Magazine, vol.24, no.2, pp.168-169, 2007.

[3] F. Dufaux, *Video Scrambling for Privacy Protection in Video Surveillance: Recent Results and Validation Framework*, SPIE, Mobile Multimedia/Image Processing, Security, and Applications 2011.

[4] K. Chinomi, N. Nitta, Y. Ito & N. Babaguchi, *PriSurv: Privacy Protected Video Surveillance System Using Adaptive Visual Abstraction*, Advances in Multimedia Modeling, Lecture Notes in Computer Science, vol.4903, pp.144-154, 2008.

[5] Y. Chang, R. Yan, D. Chen & J. Yang, *People Identification with Limited Labels in Privacy-Protected Video*, IEEE Int. Conf. on Multimedia and Expo, pp.1005-1008, 2006.

[6] Y. Iwashita & R. Kurazume, Person identification from human walking sequences using affine moment invariants, Proc. IEEE Int. Conf. Robotics and Automation, pp.436-441, 2009.

[7] E. Rahtu, M. Salo, J. Heikkila & J. Flusser, Generalized affine moment invariants for object recognition, Proc. IEEE Int. Conf. on Pattern Recognition, pp.634-637, 2006.

[8] B. Babenko and M. Belongie, *Robust Object Tracking with Online Multiple Instance Learning*, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2011.

[9] J. Han & B. Bhanu: Individual recognition using gait energy image, Trans. on Pattern Analysis and Machine Intelligence, Vol. 28, No. 2, pp.316-322, 2006.

[10] E. Zhang, Y. Zhao & W. Xiong, Active energy image plus 2dlpp for gait recognition, Signal Processing, vol.90, no.7, pp.2295-2302, 2010.

[11] J. Flusser, T. Suk & B. Zitova, Moments and Moment Invariants in Pattern Recognition, Wiley & Sons Ltd., 2009.