2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing

A Method for Automatic Detection of Crimes for Public Security by Using Motion Analysis

Koichiro Goya, Xiaoxue Zhang, Kouki Kitayama
Graduate School of Engineering, Faculty of Engineering,
University of the Ryukyus
1 Senbaru, Nishihara, Okinawa JAPAN
{goya/zhang/rance}@engr.ie.u-ryukyu.ac.jp

Itaru NAGAYAMA
Department of Information Engineering, Faculty of Engineering,
University of the Ryukyus
1 Senbaru, Nishihara, Okinawa JAPAN
nagayama@ie.u-ryukyu.ac.jp

Abstract

In this paper, an automated video surveillance for crime scene detection using statistical characteristics is presented. The system is named Public Safety System(PSS). If the scene shows some peculiar situation such as purse-snatching, kid napping and fighting on the street, the PSS recognize the situation and automatically report to agency. Localization of moving targets in the scene and human behavior estimation are key processes of the proposed method. Three motion characteristics are determined from video stream: distance between objects, moving velocity of objects and area of objects. These characteristic are used to determine human behavior. Using these three metrics as a feature vector, the system classify video streams into criminal and non-criminal scenes. We use these two kinds of action sequences for the training data set. After constructing the classifier, we use test sequences those are continuous video stream of human behavior consists several actions in succession. The experimental results show the method is useful to detect criminal scene by the discrimination of human behavior.

1. Introduction

Visual surveillance in machine understanding has been investigated worldwide during last decades. Human-behavior analysis and understanding from video imagery is one of the most active research fields.

In recent years, street crimes such as purse-snatching and robbery are increasing in urban areas. A "street crimes" is the generic name of crimes that take place in street, park and other public places. The street crimes have marked about 45 % of the entire crimes in Japan. For example, as a brutal street crime, mass murder case happened in Tokyo June 2008 has recorded 17 victims and surprised all over Japan. Thus, the street crimes become major social concerns.

Especially, a purse-snatching, one of the crimes, is becoming more common in urban area because of the recession in recent years. This tendency seems to be prevalent nationwide. According to the Police Agency, the number of purse snatching cases around the nation between January and April was about 7,000, which is up 14 % from a year ago. Because purse-snatching is a crime that anybody can do anywhere, especially in a empty street, it hard to predict when and where the snatching crimes will take place. This background makes investigation by police agency to be very difficult. On the other hand, CCTV(Closed Circuit TV) is a surveillance system as means to observe public spaces. One of its advantages is recording the evidence of the crimes. Another advantage is the ability to control the crimes by existence of many cameras. Although this conventional surveillance performed manually is efficient for crime prevention, it requires many human resources and is expensive. Moreover, it will be difficult to manipulate a great many of CCTV cameras in near future.

In this paper we present a method that uses the morphological information between detected objects in video sequence to classify the video scenes into purse-snatching or non-criminal. We show the proposed Public Safety System has a good performance for automatic crime detection on

978-0-7695-3762-7/09 $26.00 © 2009 IEEE
DOI 10.1109/IIH-MSP.2009.264

the street. The organization of this paper is as follows. In section 2 we describe some previous works related to this paper. Section 3 explains the algorithm for scene recognition with a detailed description of each step and some examples. Section 4 describes system implemantation and experimental results. Conclusions and future directions are outlined in section 5.

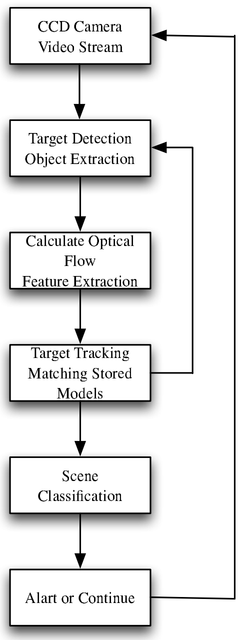

Figure 1. System Implementation

2. Previous Works

There have been several works on machine vision. An early paper on real time automated video surveillance was presented by O'Rourke and Badler[1]. Then, several works belong to model based approaches and model free approaches have been presented. Haritaoglu et.al. proposed a system which uses projection histograms of the single detected object as features to classify different actions[2]. Cucchiara et.al. presented a video surveillance system for human behavior recognition[3].Wren et.al. presented a well performing video surveillance system[6]. This is active video sensing devices with multiclass statistical models, and it can recognize objects in the indoor scene.

3. Proposed Method

The recognition of human behavior from surveillance camera is one of the most challenging problems due to the difficulties in variety of the surrounding conditions. However, real-world implementations have to be computationally inexpensive and be applicable to actual scenes where targets are small, rough and video stream is noisy. Thus, the methodology of using a object's outline for analyzing its motion is often adopted under these conditions. In this paper, we analyze continuous human behavior by automatically segmenting the moving objects in video stream. For example, the situation of purse-snatching is actually movement of relevant people and objects. The scenario to be solved for motion understanding and computer vision is commonly consists of several processing steps: video stream acquisition, object detection, object segmentation, feature extraction and classification. The implemented video surveillance system named PSS consists of four sequential blocks: acquisition, object detection/segmentation, feature extraction and scene classification as shown in Fig.1. In particular, target detection block requires optical flow for automatic extraction of the velocity of interested moving objects.

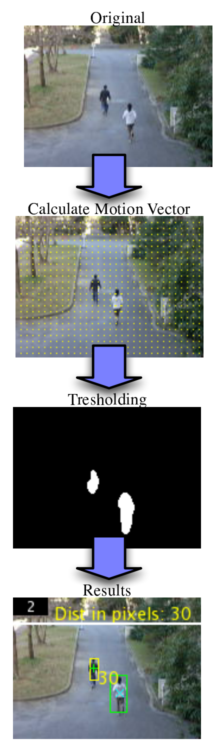

Figure 2. Purse-snatching detection

3.1. Moving Object Detection

The first step of the human motion analysis is detection of moving objects from a video stream. There are some fundamental approaches to moving objects extraction: background subtraction, temporal differencing, and optical flow. Background subtraction does the most complete extraction of all relevant pixels, however it is very sensitive to dynamic scene changes due to the lighting. Temporal differencing is flexible to variety of the video capturing condition, but it does poor work to detect all relevant pixels. Optical flow can detect moving objects and its vector fairly in spite of the existence of camera's motion, but it require complicated computation. This system uses the optical flow and the background subtraction to detect the moving targets and to compute the average velocity of each moving object in video stream. The average velocity is calculated from vector information obtained by optical flow. The examples are illustrated in Fig.2.

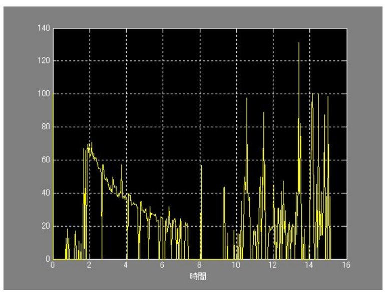

Figure 3. Distance

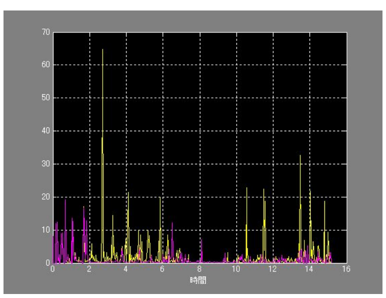

Figure 4. Velocity

The algorithm for motion analysis uses the distance between people, denoted by x, the moving speed of each object,y, and the pixels of objects, z. These three numbers form a feature vector(x, y, z). Each frames of a video sequence is represented by a feature vector (x, y, z). We define two classes, a purse-snatching class and non-snatching class. Training vectors for both classes are chosen from a set of video sequences. During a continuous video sequences, x, y and z traverses a series of maxima and minima. Fig.3,4 and 5 show x, y and z respectively for a video sequence illustrated in Fig.6.

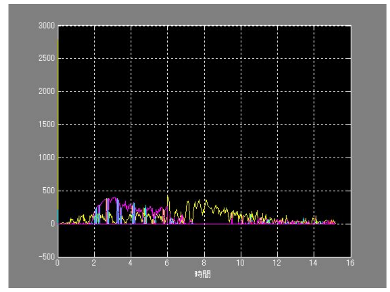

Figure 5. Area

3.2 Target Tracking

The noise reduction processing is to clean up anomalies in the detected targets. This is done by a morphological dilation followed by erosion. This step removes any small holes in the object and smoothes any interlacing anomalies. These dilation and erosion effectively preserve small features such as thin leg or body segments in the video frames. After these objects are cleaned up, their outline are extracted by using rectangular segmentation. The centroid of the target image is calculated in Eq.1.

$$P_{ck} = \frac{1}{N_k}(\sum_{i=1}^{N_k}(P_i)), Q_{ck} = \frac{1}{N_k}(\sum_{i=1}^{N_k}(Q_i)) \quad (1)$$

where (P_{ck}, Q_{ck}) is the centroid of target k, N_k is the number of pixels within the target k, and (P_{ik}, Q_{ik}) is a pixel on the extracted area of target k. The distance d_{ik} between the centroid (P_{ck}, Q_{ck}) and pixels (P_{ik}, Q_{ik}) on the target's outline is calculated as Euclidean distance. Also the distance d_{ckl} between centroids (P_{ck}, Q_{ck}) and (P_{cl}, Q_{cl}) is calculated. There are some moving targets, then distances of each pair of centroids of targets are continuously computed for tracking their geometrical relation. The area of each moving target is used to distinguish its property that is human body or not. Although one of those distances between centroids is decreasing, it may be not shorter than personal area for ordinary situation. An extraordinary situation is detected by finding the realization of the function

$$d_{ckl} \leq d_{ik} + d_{il} \quad (2)$$

Motion analysis has been tried on a set of video sequences of walking people at outdoor on a daytime. There are approximately 20 video sequences in each category. Pixels on target ranging from 40 to 500 caused by the distance from camera to the targets. The motion analysis is performed on Dell PowerEdge server with dual Xeon 2.4GHz processors.

Figure 6. Frames 1 to 16 of purse-snatching.

4. Analysis and Recognition

The recognition process is performed by using nearest neighbor classifier for two classes as shown in Fig.8 and Fig.9. Video sequences are used to train the system. Half of them is purse-snatching scenes and the other half is not purse-snatching. Three motion characteristics are determined from video stream: distance between objects, average moving velocity of objects and area of objects. These characteristic are used to classify the scenes, because they are different from each other. The three-element test feature vector (x_i,y_i,z_i) is compared with the training feature vectors from each class. The algorithm computes Euclidean distance measures D between the feature vectors for each video sequence of a test sequence and the training feature vectors of the snatching and non-snatching classes respectively. Let D_i be the Euclidean distance defined as:

$$D_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2 + (z_i - z_c)^2} \qquad (3)$$

where (x_i,y_i,z_i) defined as feature vector and (x_c,y_c,z_c) the position of the centroid of each class. A video sequence is classified as a scene of purse snatching if the minima of the distance measure set D is less than the criteria D_p. We are basically looking for combinations of criteria that characterize the difference between classes.

4.1. Experimental Results

Figure 6 shows a video frames of purse-snatching on a street. Although a person is walking on street from this side to over there, a criminal chases the walking person to snatch her/his bag from backward. After purse-snatching is done, a criminal run away from the place. Figure 7 shows a video frames of same place. Although a person follows another one from backward, he passed by without anything. Both video sequences should be classified correctly.

In addition to that, the situation that a person is walking

Figure 7. Frames 1 to 16 of passing by.

Table 1. Results for test sequences

Walking Direction	snatching	no snatching
To over there	80%	80%
To this side	80%	60%

on street from over there to this side is also experimented as same as above. These data were acquired in real time from a video camera with 120x160 pixels sequence. This value is constant in these experiments. Video sequences are obtained by a fixed CCD video camera working at 30 frames per second. The test sequences range in time from 25 to 35 seconds. The system detects every moving targets in video stream. If there are some walking or running person that consists of adequate pixels area, the system segments the interesting targets from background. After the segmentation, the resulting image targets are focused on and are continuously tracked. The distances between these targets are calculated. Moving speed and area of each target are also calculated. A sudden change of target's velocity and moving direction are carefully investigated. These metrics are continuously monitored. If the distance between two moving objects is decreasing, it should be attract more attention. Then extracted feature vectors are classified using the nearest neighbor classifier. The nearest neighbor classifier assigns the feature vector$(x_k, y_k, z_k)_i$ to the same class C_k,(where $k \in 1, 2$) as the feature vectors nearest to the centroid of each class in the feature space. The implemented system has been tested on 20 sequences of video stream. Half of them is purse-snatching scenes and the other half is of non-snatching scenes. Table 1 shows the results that have been obtained on the test sequences. The results demonstrate the efficiency of the method with respect to detection of purse-snatching scene. They are expressed in terms of classification rate for different walking directions.

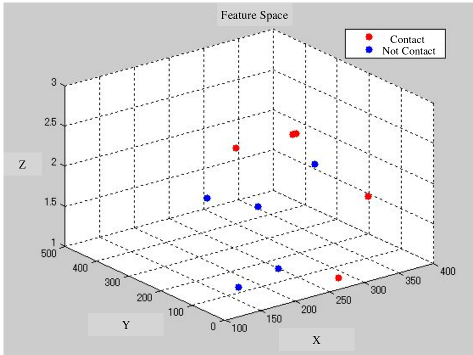

Figure 8. Position in feature space.

5. Conclusions

We have presented a method for automatic detection and recognition of continuous human activity. The human activity sequence consists of visual actions that are performed in succession. This paper presented a video surveillance for scene classification based on object segmentation and its movement. In general, real-world implementations have to be computationally inexpensive and be applicable to real scenes where target objects are small and video stream is noisy. The using a object's outline for analyzing its motion is often adopted under these conditions. Although the system is limited to the longitudinal viewpoint of the object, we have tested it with sequences in which the camera does not view the subject from a perfect longitudinal view. Different directions of the object's movement on opposite line of the camera have been tested. Experimental tests for human behavior estimation have been execute considering three different feature series: the Euclidean distance of the center position from segmented objects, the average speed of moving objects and the pixels of target objects. Good performances have been obtained for both cases, purse-snatching or not, with simple features. In the near future, the presented method will be applied to detect more complicated human behaviors of crimes such as quarrel, fighting, kid napping and so on.

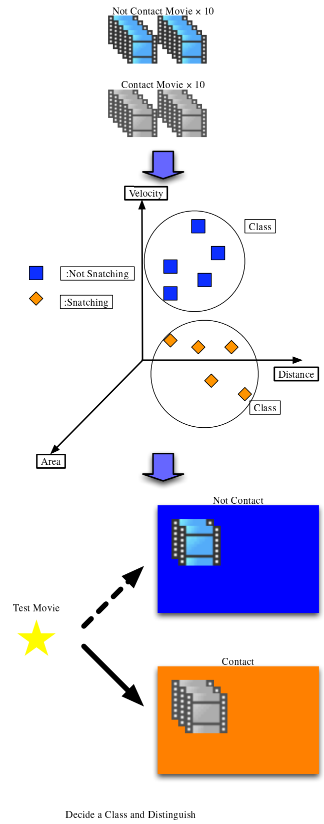

Figure 9. Scene classification by k-NN.

References

[1] J. O' Rourke and N. Badler, "Model-based imageanalysis of human motion using constraint propagation",IEEE Trans. on Pattern Anal. and Mach. Intell.,2(6),pp.522-536, November(1980)

[2] I.Haritaoglu, D.Harwood and L.S.Davis, "W4: Realtime surveillance of people and their actions",IEEE Trans. on Pattern Anal. and Mach. Intell.22(8),pp.809-830, August (2000)

[3] R. Cucchiara, C. Grana, A. Prati, R. Mezzani,"Probabilistic posture classification for human-behavioranalysis" , IEEE Trans. On System, Man and Cybernetics, 35(1), January(2005)

[4] A.F.Bobick and J.W.davis,"The recognition of human movement using temporal templates",IEEE Trans. Pattern Anal. Mach. Intell. 23(3),pp.257-267, March(2001)

[5] Hironobu FUJIYOSI and Takeo KANADE,"Layered Detection for Multiple Overlapping Object",IEICE Trans. Info. & Syst., Vol.E87-D,12 December 2004

[6] C.R.Wren, A.Azarbayejani, T.Darrel and A.P.Pentland,"Pfinder: real-time tracking of human body",IEEE Trans. Pattern Anal. Mach. Intell. 19(7),pp.780-785, July(1997)

[7] G. Loy, L.Fletcher, N.Apostoloff, and A.Zelinsky,"An adaptive fusion architecture for taget tracking",Proc. 5th IEEE International Conference on Automatic Face and Gesture Recognition,pp.261-264 May (2002)

[8] A.Mohan, C.Papageorgiou and T.Poggio,"Example-based object detection in images by components",IEEE Trans. Pattern Anal. Mach. Intell. 23(4),pp.349-361, April(2001)