2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing

A Markov Random Walk Model for Loitering People Detection

Thi Thi Zin, Pyke Tin and Takashi Toriu
Graduate School of Engineering
Osaka City University, Osaka, Japan
{ thithi; pyketin }@ip.info.eng.osaka-cu.ac.jp,
toriu@info.eng.osaka-cu.ac.jp

Hiromitsu Hama
R&D Center of 3G Search Engine, Incubator
Research Center for Industry Innovation,
Osaka City University, Osaka, Japan
hama@ado.osaka-cu.ac.jp

***Abstract*— Today video surveillance systems are widely used in public spaces, such as train stations or airports, to enhance security. In order to observe large and complex facilities a huge amount of cameras is required. These create a massive amount of data to be analyzed. It is therefore crucial to support human security staff with automatic surveillance applications, which will create an alert if security relevant events are detected. This way video surveillance could be used to prevent potentially dangerous situations, instead of just being used as forensic instrument, to analyze an event after it happened. In this treatise we present a surveillance system which supports human operators, by automatically detecting loitering people. Usually, loitering human behavior often leads to abnormal situations, like suspected drug-dealing activity, bank robbery, and pickpocket, etc. Thus, the problem of loitering detection in image sequences involving situations with multiple objects is studied based two dimensional Markov random walks in which both motion and appearance features describing the movements of a varying number of objects as well as their entries and exits are used. To obtain efficient and compact representations we encode the spatiotemporal information of intra-inter trajectory contexts into the transition matrix of a Markov Random Walk, and then extract its stationary distribution and boundary crossing probabilities as final detection criteria. The model is also made less sensitive to uninteresting objects occluding the region of interest by integration out their effect on the observation probabilities. The resulting system is tested on the real life dataset scenarios giving 95% performance results.**

Keywords- *loitering, video surveillance, human behavior, Markov random walk model, trajectory*

I. INTRODUCTION

The problem of loitering is as old as the civilized societies, although its contours have changed over the years. The poor asking for money and the homeless seeking a place to sleep have always troubled residents, particularly those living in the close quarters of a city. Charitable and civil law efforts in the societies around the globe ease some of the problems, but for the aggressively idle or the overtly unpleasant, the usual response has been the criminal law. Although loitering and vagrancy statutes typically specify at least some prohibited acts, usually the conduct is not itself harmful. Instead, the prohibited conduct is often described in a way that correlates to the behavior of young men who are most likely to associate with gangs, whether as members or otherwise. Singling out this group is both predictable and rational, since its members commit a disproportionate amount of crime in general, and the street-level crimes of theft, drug possession and sale, assault, and robbery in particular. In most contexts the generalized threat of future criminality is not a basis for intervention; the authorities must either wait until a crime is attempted or engage in costly monitoring or undercover work to intervene to prevent the crimes. Much of the crime engaged in by gang members, however, stubbornly defies these efforts. The opportunistic loiterer who is selling drugs, committing small thefts, spray painting graffiti, or intimidating pedestrians is often in no hurry. He can wait until the police pass by and wait for another customer or another victim, making monitoring extremely expensive. Undercover work is also difficult, in part because infiltration of groups is risky and in part because some of the crimes involve little advanced planning. Finally, waiting for the crimes to occur and then investigating is a frustrating and often fruitless enterprise. Victims are reluctant to testify against gang members for all of the obvious reasons. More importantly, residents will often simply change their behavior-they will not walk in those city blocks, will not go out after dark, or in extreme cases, will move out of the neighborhood entirely.

Given these difficulties, the power to disperse a troublesome group or loitering people is to create a vision-based system to monitor inner-city areas for suspected drug-dealing and potential crime oriented loitering people activities. This is a major problem for Metro Transits such as bus stops, train stations and airports, because drug-dealing activity makes their bus stops dangerous and discourages citizens from using them. According to officials at the Metro Transits, drug dealers will typically loiter around bus stops while their customers come in on buses, purchase their material, and then leave on other buses. Thus, suspected drug-dealing behavior can be detected by monitoring a bus stop for people loitering about longer than would be necessary to catch a bus. If we were to monitor a bus stop for an extended period of time, we would notice that people either just walk by or wait for the bus for a period of time that rarely exceeds 0.5h after which they board a bus. Again, one of the main characteristics of drug-dealing behavior is the presence of the dealer for extended periods of time without boarding any bus. Moreover, individuals loitering near access doors may be looking to "tailgate" an authorized member of staff in order to gain access to the premises or to a restricted area. They might alternatively be tampering with

978-0-7695-4222-5/10 $26.00 © 2010 IEEE
DOI 10.1109/IIHMSP.2010.172

security equipment in order to avoid detection. Loitering youths may vandalize equipment or buildings, such as spraying graffiti or tags. They may alternatively be drinking or taking illegal drugs, or trying to sell drugs to passers-by.

Recently, some researchers [1-5] are leading studies to detect these situations of loitering in the context of inner-city bus stops, mainly to detect drug-dealers. However, the existing loitering systems have not yet the ability to match a new image of a pedestrian to a pedestrian that it has already seen, essentially a system that can pick someone it has seen before out of a suspect lineup. However, in this lineup, everyone is not necessarily facing the same direction, or in front of the same background, making automated recognition difficult. On the other hand, previous approaches have reported many difficulties including shadows, occlusions, non rigid targets, and varying lighting conditions. Other problems lie in the classification of pedestrians seen such as discerning the volume of data required to discriminate between varying numbers of pedestrians, what features to use to differentiate individuals, and determining when a given piece of data is no longer useful for classification. Therefore the action recognition algorithms [6-8], which aimed to achieve the robustness to viewpoint change based on geometric reconstruction, are doomed due to their reliance on exact knowledge of the object contours or multiple view geometries, which are prone to errors in unconstrained videos.

In this paper, we propose a Markov random walk model by using the appearance and motion features based on the spatio-temporal co-occurrence and distribution of the trajectories. The Markov random walks have been successfully applied in queuing [9], dams [10] and image search engines [11], we can derive compact and efficient representations of context based on their stationary distributions and passage crossing probabilities. Employing the rule based algorithms [12] to fuse feature channels we validate our proposed framework over a realistic action database and demonstrate its superiority over the state-of-the-art.

The rest of the paper is organized as follows. Section 2 gives a brief overview of how the proposed loitering detection system functions. Section 3 describes in depth the Markov random walk module of detection process, and Section 4 discusses experimental results, and Section 5 presents the conclusions.

II. THE SYSTEM OVERVIEW

The system described here can be divided into three functional components. The first component is the pedestrian segmentation. This is the part of the system that goes through the video and extracts pictures of different pedestrians as well as rough tracking data to pass on to be classified and analyzed. The second component is the correlation module. This module extracts appearance and motion features in addition to the trajectories extractions from the images passed from the segmentation module and classifies them as either one of the pedestrians seen before or a completely new pedestrian. In the third component, each trajectory or the entire video is considered as a Markov random walk system by using the motion and appearance features such as velocities and the center of gravities as underlying probability distribution for a random walk. We can then construct the fundamental equation of the random walk model and derive compact and efficient representations of context based on their stationary distributions and boundary crossing probabilities for the relevant regions extracted in all field of views (FOV) with a foreground segmentation module. By using a suitable threshold, for example the boundary crossing probability must be less than Th_B (say 50%) and the sum of the cumulative stationary distributions must be higher than Th_S (say 90%) we can say the potential loitering person will be hanging around in the detected region long enough to be classified as actual loitering person. In this way, the overview of our proposed loitering people detection system can be described as in Fig.1.

III. MARKOV RANDOM WALK MODEL

In this section we shall describe the Markov random walk model which is to be employed in our proposed system. It is worth to note that the point-level context in image processing characterizes mainly what part of the video, namely what object components appear in the videos. For action recognition, however, the dynamic properties of these object components are more essential in characterizing the actions, e.g. for the action of loitering or handling an object etc. The Markov random walk and Markov chain are the most powerful tools for modeling the dynamic properties of a system. Its merit mainly lies at its capability in representing directed causal and probabilistic relations. The Markov stationary distribution and passage probabilities, associated with an ergodic Markov chain, offer a compact and effective representation for a dynamic system.

We will now consider a random walk in the plane, within the rectangle [1; M-1] x [1; N-1], where $M \geq 2$ and $N \geq 2$ are integers, as shown in Fig.2. We assume that the walker starts at an inner point (m, n). The walker moves to the neighboring points (m+1, n), (m-1, n), (m, n+1) or (m, n-1) with probabilities p_1, p_2, p_3 and p_4, which are derived the motion features in our case respectively. The symbols ∇ and ◊ are used for the upper and lower boundary points. Sometimes, they are called impenetrable barriers.

Let $u(m, n)$ be the probability that the random walk is ending at the upper horizontal part of the boundary, starting at (m, n). Then u will satisfy the following partial difference equation:

$$\begin{aligned} u(m,n) = p_1\, u(m+1,n) + p_2\, u(m+1,n) \\ + p_3\, u(m,n+1) + p_4\, u(m,n-1). \end{aligned} \tag{1}$$

Obviously we have the boundary conditions:

$$\begin{aligned} &u(0,n) = u(M,n) = 0, && n = 1 : N-1, \\ &u(m,0) = 0, \text{ and } u(m,N) = 1, && m = 1 : M-1. \end{aligned}$$

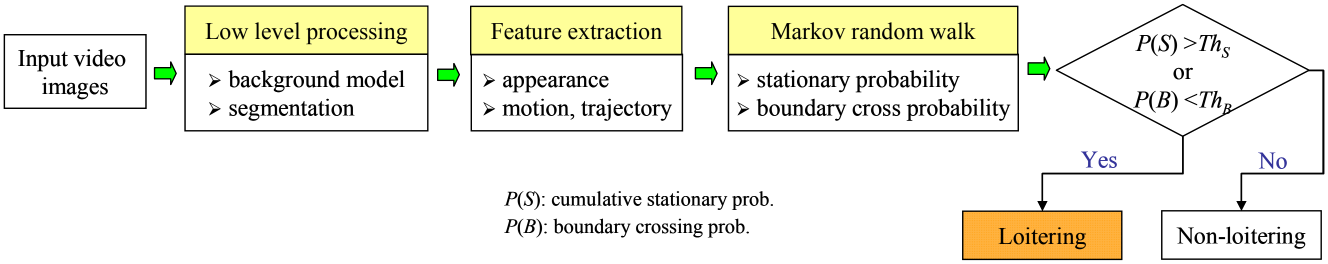

Figure 1. System overview.

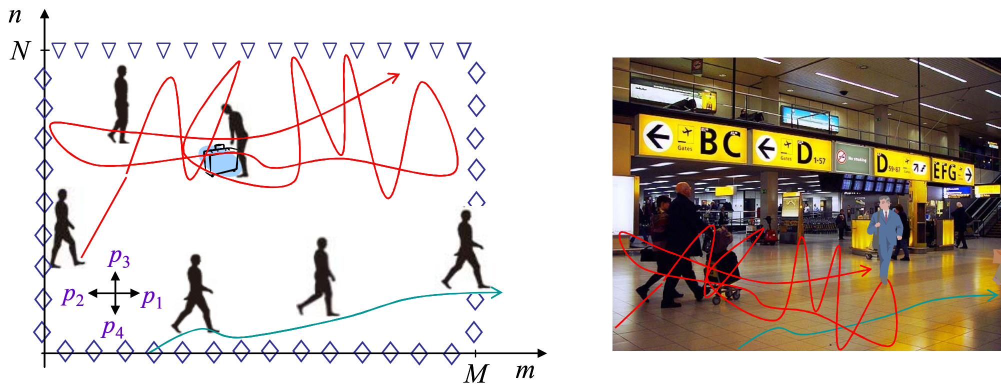

Figure 2. Markov random walk in plane.

By solving Eq.(1), we obtain the probability of crossing boundary $u(m, n)$ which we will use for detection process with a suitable threshold. In similar ways, we can derive the stationary distribution of the random walk model.

Let the stationary distribution be denoted by

$$\pi = [\pi_1 \ \ \pi_2 \ \ \ldots \pi_k].$$

We then have the cumulative stationary probability $P(S)$ and the boundary crossing probability $P(B)$ as follows:

$$P(S) = \sum \pi_j, \quad P(B) = u(m,n).$$

Thus, the criteria for detecting loitering is

$$P(S) > Th_S \ \text{ or } \ P(B) < Th_B \,. \qquad (2)$$

The visual concept of this criterion can be illustrated as shown in Fig. 3.

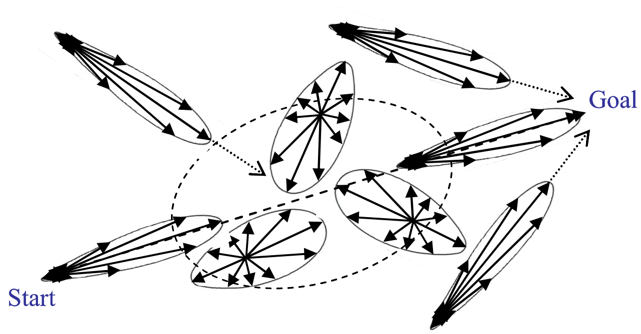

Figure 3. Visual concept of loitering.

In addition, to detect loitering from the classes within the database, the time stamps passed to this module with each pedestrian instance are used. The time difference between the first recorded frame the pedestrian was detected and the last recorded frame the pedestrian was detected is checked for each class in the database. If this time difference is greater than a threshold value, the class is then checked for gaps in detection greater than predetermined time. If no gaps are found, the pedestrian is declared to be loitering.

IV. Experimental Results

In this section, we systematically evaluate the effectiveness of our proposed method on the datasets of the PETS2007. Since our approach is targeted mainly on the human loitering behavior detection, we have done the person tracking manually. Loitering persons were recognized without any misses. According to the definition provided for the PETS2007 challenge a person is loitering if he stays in the field of view for at least 60 sec. This task can be solved rather easily and accurately by considering the stationary and boundary crossing probability as described in our Markov random walk model for every tracked object if it appears for the first time. We define the loitering zone as the areas within the camera view. We also use 0.95 as the threshold for stationary probability and 0.5 for boundary crossing probability. The first threshold 0.95 means that the probability of the potential loitering person within the loitering zone is almost certain in specified time duration. On the other hand the use of second threshold value 0.5 states that the probability of the person crossing the zone boundary is very low compare to the corresponding probability of normal people. These concepts have made our experiment to produce more realistic results. Fig.4 shows one example of detected loitering person among other normal pedestrians on a sequence of images using our proposed methods. Some examples of PETS2007 images used in our experiments are shown in Fig. 5.

V. CONCLUSIONS

We have proposed and investigated a Markov random walk model that can robustly detect loitering individuals in any outdoor public place. The system is divided into three main components. The segmentation module processes the video and extracts images of individual pedestrians by using background subtraction and blob tracking. These images, along with their most probable backgrounds, time stamp, and tracking number are then passed to the feature extraction module proceeding towards the Markov random walk modeling module for detecting loitering people. To determine if a given class is loitering, the time stamps associated with it are analyzed by using stationary and boundary crossing probabilities. This model has been evaluated on realistic action and event recognition databases, and has shown superior performance over the state-of-the-art features and algorithms.

ACKNOWLEDGMENT

This work is partially supported by KAKENHI 21500176 Grant-in-Aid for Scientific Research(C).

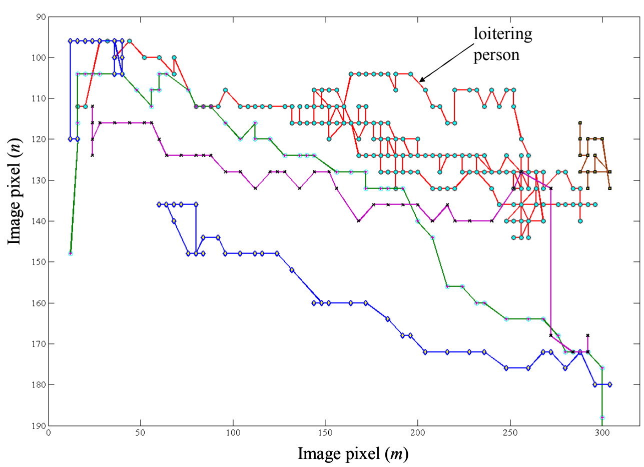

Figure 4. Example of loitering people detection from one image sequence.

Figure 5. Example frames from PETS2007 dataset.

REFERENCES

[1] N. Bird, O. Masoud, N. Pananikolopoulos, and A. Isaacs, “Detection of loitering individuals in public transportation areas,” *IEEE Trans on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 167-177, Jun. 2005.

[2] Hakan Ardo and Kalle Astrom, “Multi sensor loitering detetion using online viterbi,” *Proc. of IEEE Intl. Workshop on Performance Evaluation of Tracking and Surveillance* (*PETS2007*), Rio de Janeiro, Brazil, Oct. 2007.

[3] G. Gasser, N. Bird, O. Masoud and N. Papanikolopoulos, “Human activities monitoring at bus stops,” in *Proc. IEEE Int. Conf. Robotics andAutomation*, New Orleans, LA, Apr. 2004, vol. 1, pp. 90–95.

[4] D. Arsi´c, B. Schuller, and G. Rigoll, “Suspicious behavior detection in public transport by fusion of low-level video descriptors,” *Proc. of 8th Intl. Conf. on Multimedia and Expo* (*ICME2007*), Beijing, China, Jun. 2007.

[5] Chung-Hsien Huang, Yi-Ta Wu, Ming-Yu Shih, “Unsupervised Pedestrian Re-identification for Loitering Detection,” *Proc. on the 3rd Pacific-Rim Symposium on Image and Video Technology* (*PSIVT 2009*), Tokyo, Japan, pp. 771-783, Jan. 13-16, 2009.

[6] A. Yilmaz and M. Shah, “Actions Sketch: A Novel Action Representation,” *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (*CVPR’05*), vol.1, pp.984-989, 2005.

[7] A. Yilmaz and M. Shah, “Recognizing Human Actions in Videos Acquired by Uncalibrated Moving Cameras,” *Proc of 10th IEEE Intl. Conf. on Computer Vision* (*ICCV’05*), Beijing, China, vol. 1, pp.150-157, Oct. 15-21, 2005.

[8] Arnold Wiliem, Vamsi Madasu, Wageeh Boles, and Prasad Yarlagadda, “Detecting Uncommon Trajectories,” *Proc. of Digital Image Computing: Techniques and Applications* (*DICTA*), Canberra, Australia, pp. 398-404, Dec. 1-3, 2008.

[9] J. Gani and Pyke Tin, “On a class of continuous-time Markov processes,” *J. Appl. Prob.*, vol. 22, pp.804-815, 1985.

[10] R.N.Phatarfod and K.V.Mardia, “Some Results for Dams with Markovian Inputs,” *J.App.Prob.*, vol.10, pp166-180, 1973.

[11] Hiromitsu Hama, Thi Thi Zin and Pyke Tin, “Ranking System for Image Database Using Special Type of Markov Chain,” *Proc. of Digital Image Computing: Techniques and Applications (DICTA2008)*, Canberra, Australia, pp. 556-563, De c. 1-3, 2008.

[12] Thi Thi Zin, Pyke Tin, Takashi Toriu and Hiromitsu Hama, “Dominant Color Embedded Markov Chain Model for Object Image Retrieval”, *Proc. of the 5th Intl. Conf. on Intelligent Information Hiding and Multimedia Signal Processing* (*IIH-MSP-2009*), Kyoto, Japan, pp. 186-189, Sep. 2009.