2012 IEEE International Conference on Robotics and Automation
RiverCentre, Saint Paul, Minnesota, USA
May 14-18, 2012

Laser-based Intelligent Surveillance and Abnormality Detection in Extremely Crowded Scenarios

Xuan Song, Xiaowei Shao, Quanshi Zhang, Ryosuke Shibasaki, Huijing Zhao and Hongbin Zha

***Abstract*— Abnormal activity detection plays a crucial role in surveillance applications, and a surveillance system that can perform robustly in the extremely crowded area has become an urgent need for public security. In this paper, we propose a novel laser-based system which can simultaneously perform the tracking, semantic scene learning and abnormality detection in the large and crowded environment. In our system, a novel abnormality detection model is proposed, and it considers and combines various factors that will influence human activity. Moreover, this model intensively investigate the relationship between pedestrians' social behaviors and their walking scenarios. We successfully applied the proposed system to the JR subway station of Tokyo, which can cover a 60×35m area, robustly track more than 180 targets at the same time and simultaneously perform the online semantic scene learning and abnormality detection with no human intervention.**

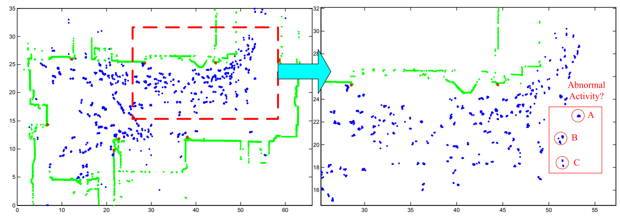


Fig. 1. How to detect abnormal activity in the extremely crowded area? This is the JR subway station of Tokyo, and the data was obtained by eight single-row laser scanners. The green points are the background, the blue ones are the foreground, and the red ones show the position of single-row laser scanners. In this case, each person is represent by several points. For more details about the experimental site, please refer [1]. Person A, B and C were walking on the closed road, how to detect their activities?

I. INTRODUCTION

The intelligent surveillance system which can cover a large and crowded public area has become an urgent need for the public security, and one of its key components is to detect abnormal behavior patterns and recognize the normal ones. However, most existing systems are usually based on camera, which can only cover a small area, and it difficult for them to work robustly in some extremely crowded scenarios, such as subway station, public square, intersection and etc (as shown in Fig.1). *Therefore, the purpose of this paper is to develop a novel intelligent system and abnormality detection model that can perform the detection of abnormal activities robustly in the large and extremely crowded scenarios.*

While the pedestrians are walking in a specific scenarios, their activities will be greatly influenced by the semantic scene knowledge (e.g., dominant paths, entry or exit, crowd flow and etc.). For instance, "persons usually walk from entrance to exit", "normal persons have to walk in the dominant paths and avoid static obstacles", "normal persons who are in a crowd flow can only follow the other people in it." A statistical scene model can provide a priori knowledge on where, when and what types of activities occur. Therefore, in this paper, *we intensively investigate the relationship between pedestrians' social behaviors and their walking scenarios, and propose a novel abnormality detection model to detect abnormal activities of persons in the large and extremely crowded environments*. Our model considers and combines various factors that will influence human activity (as shown in Fig.2), such as global semantic scene structure (paths, exit/entrance), local instantaneous crowd flow, centrifugal force among pedestrians. Lastly, we apply this model into an "adaptive tracking-learning loop" and *develop an intelligent surveillance system which can simultaneously perform the tracking, semantic scene learning and abnormality detection in a fully online way.*

The main contributions of this paper can be summarized as follows: (1) We propose a novel abnormality detection model which intensively investigate the relationship between pedestrians' social behaviors and their walking scenarios. (2) We develop a unified framework that couples the tracking, semantic scene learning and abnormality detection, and make them supplement each other in it. (3) We firstly apply an online system that can robustly track more than 180 targets at the same time and perform robustly abnormality detection to a real scene (JR subway station of Tokyo).

The remainder of this paper is structured as follows: In the following section, related work is briefly reviewed. Section III provides an overview of the proposed system. Section IV and V provide the details about the abnormality detection model and "tracking-learning" module. Experiments and results are presented in Section VI and the paper is finally summarized in Section VII .

II. RELATED WORK

Abnormality detection is an active area of research over these years, and an in-depth review of its literature can be found in a recent survey by Chandola *et al.* [2]. Traditional

This work was supported by a Grant-in-Aid for Young Scientists (23700192) and "Strategic Project to Support the Formation of Research Bases at Private Universities": Matching Fund Subsidy by Japan's Ministry of Education, Culture, Sports, Science, and Technology (MEST). This work was also partially funded by East Japan Railway Company and Microsoft Research.

Xuan Song, Xiaowei Shao, Quanshi Zhang and Ryosuke Shibasaki are with the Center for Spatial Information Science, The University of Tokyo, Japan. E-mail:{songxuan,shaoxw,zqs1022,shiba}@csis.u-tokyo.ac.jp.

Huijing Zhao and Hongbin Zha are with the Key Laboratory of Machine Perception (MoE), Peking University, China. E-mail:{zhaohj,zha}@cis.pku.edu.cn.

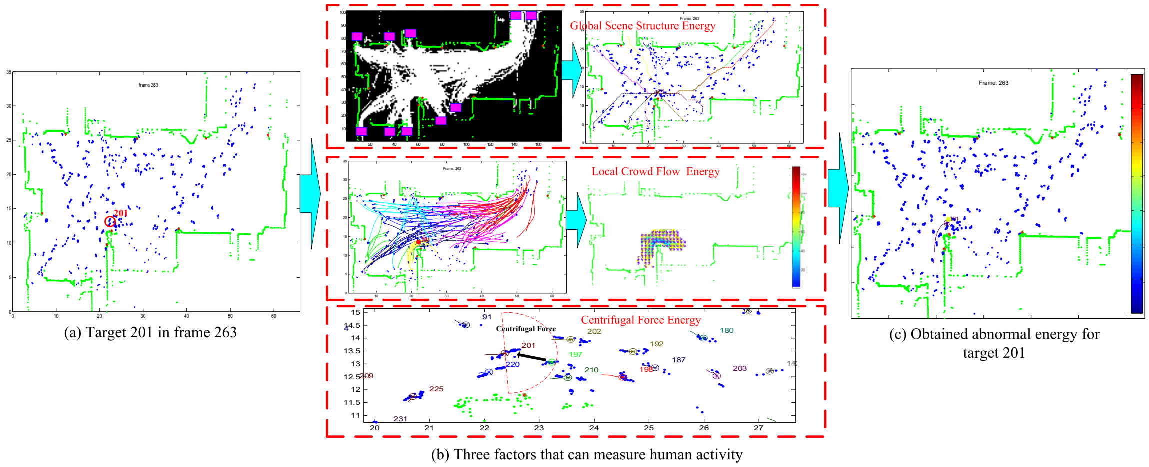


Fig. 2. Overview of the abnormality detection model. While the normal pedestrians are walking in the large and high density scene (such as person 201 in Fig.a), their activities should follow three principles (Fig.b): (1) Global scene structure: person should consider the scene structure, move from entrance to exit, walk on the dominant paths, avoid obstacles and find the shortest path. (2) Local crowd flow: persons who are in a specific crowd flow have to follow other persons in it. (3) Centrifugal force: persons usually want to keep a comfortable distance from others. Based on the three principles, our model compute the *abnormal energy* of persons to measure their activities (as shown in Fig.c, the color circle shows the value of abnormal energy).

abnormality detection approaches usually need to pre-define the known priori behavior classes (normal activities vs abnormal ones), and utilized the supervised learning model to detect the abnormal ones. Representative publications include [3], [4], [5], [6], [7]. On the other hand, researchers also propose some methods [8], [9], [10], [11], [12], [13], [14] to detect abnormal activity in an unsupervised manner. These methods are usually based on the tracking. While the trajectories are collected by tracking, the clustering is performed on these trajectories and some small clusters are labeled as abnormal. More recently, some "Sparse Coding" based methods [15], [16] are proposed to detect abnormal activity. These methods are usually based on an intuition that usual events are more likely to be reconstructible from an event dictionary, whereas unusual events are not.

However, few of existing works consider the relationship between pedestrians' social behaviors and their walking scenarios. Recently, some methods [17], [18] utilize crowd flow and semantic scene knowledge to detect abnormal activity and obtained good results. But these methods can be only applied for some simple scene (e.g. single sink/source, single crowd flow). Different to these methods, in this paper, we propose a novel abnormality detection model which intensively investigate the relationship between pedestrians' social behaviors and their walking scenarios.

III. System Overview

The overall abnormality detection system is illustrated in Fig.3, and it mainly contains four module: sensor fusion module, tracking module, learning module and abnormality detection module.

In the sensor fusion module, a number of single row laser scanners are exploited, so that a quite large area can be covered, while occlusions could be solved to some extent. A time server is utilized to deal with time synchronization problem between different sensors. For registration, the laser scans keep a degree of overlay between each others, and several control points in a box are utilized for computing the transformation matrix. For more details about this part, please refer [1].

The tracking module utilizes the independent particle filter-based tracker to perform the tracking. Once the tracking results are obtained, the global scene structure and local crowd flow can be online learned with these tracked trajectories. Meanwhile, the learned scene structure and crowd flow in turn assist in the tracking module and improve the tracking results. Then, the tracking results, learned scene structure and crowd flow are combined to compute the global scene structure energy, local crowd flow energy and centrifugal force energy for each person. Lastly, our system outputs abnormal activities with the help of these energy measurements.

IV. Abnormality Detection Model

A. Model Overview

We begin by introducing notations to formulize our problem. At time t, pedestrian i is represented by $\mathbf{x}_{i,t} = (\mathbf{p}_{i,t}, \mathbf{v}_{i,t})$, where $\mathbf{p}_{i,t} = (x, y)$ denotes its 2D position on the ground plane and $\mathbf{v}_{i,t} = (v_x, v_y)$ is its velocity vector at time t. We assign the abnormal energy $E(\mathbf{v}_{i,t}, \mathbf{p}_{i,t})$ for each person to measure its activity, and the higher energy E, its activity is more likely to be the abnormal one.

While the normal pedestrians are walking in the large and crowded environment, their movements should follow three principles (as shown in Fig.2): (1) *Global scene structure*. A person usually plans to go to a specific exit of the scene, walks on the common road, avoids the obstacles and finds the shortest and comfortable path. (2) *Local instantaneous crowd flow*. At some specific time, some local areas will be quite crowded and become a crowd flow. A normal person

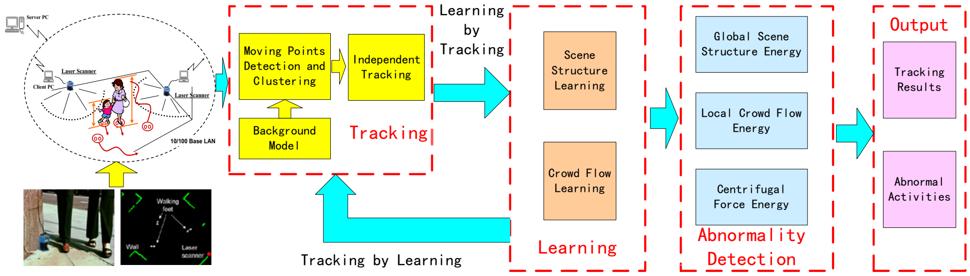


Fig. 3. Overview of the proposed system.

in a particular crowd flow will be greatly influenced by it because he must follow other persons in it. For instance, as shown in Fig.5(b), at a specific time in a subway station, a large number of persons were just getting off from a train and walking together to catch another train, which were becoming a crowd flow. (3) *Centrifugal force of pedestrians*. The activity of a pedestrian is also influenced by the *centrifugal force* [19] from its neighboring persons, as he wants to keep a comfortable distance from others, and he will feel increasing discomfort as he gets closer to a stranger. Obviously, if some persons violate these principles greatly, their activities should be the abnormal ones.

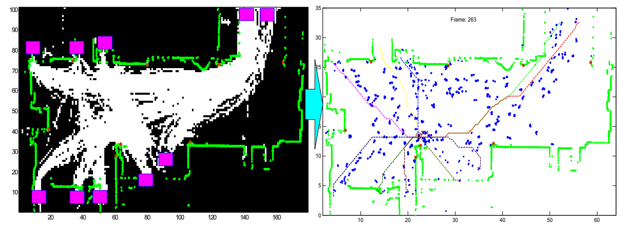
Therefore, the abnormal energy should takes into account these factors and can be computed by:

$$\begin{aligned} E(\mathbf{v}_{i,t}, \mathbf{p}_{i,t}) = \beta_{i,t} E_{global}(\mathbf{v}_{i,t}, \mathbf{p}_{i,t}) + \\ \alpha_{i,t} E_{local}(\mathbf{v}_{i,t}, \mathbf{p}_{i,t}) + (1-\alpha_{i,t}) F_{cent}(\mathbf{v}_{i,t}, \mathbf{p}_{i,t}), \end{aligned} \tag{1}$$

where E_{global} is the global scene structure energy, E_{local} the local crowd flow energy and F_{cent} the centrifugal force energy of pedestrians, $\beta_{i,t}$ and $\alpha_{i,t}$ is the control parameters, where $\alpha_{i,t} \in (0,1)$ is utilized to control the influence of E_{local} and F_{cent}, which is depend on the density of crowd flow $S_p(t)$, where $\mathbf{x}_{i,t} \in S_p(t)$. It is very easy to be understood: while the density of some local area is quite low, there would be little crowd flow or the persons' number in this crowd flow are small, and the pedestrians' activity will be greatly influenced by its nearby persons, not by the crowd flow. In contrast, the activity of persons will be more influenced by the crowd flow while the local area density becomes specially high. In the next subsections, we will provide the details about how to compute these energy functions.

B. Global Scene Structure Energy and Local Crowd Flow Energy

Given the current position $\mathbf{p}_{i,t}$ of pedestrian i by tracking, online learned scene structure map and N exits/entrances (as shown in Fig.4-a), it is easy for us to obtain N planned trajectories $\{L^l_{i,t}(x,y)\}^N_{l=1}$ for pedestrian i at time t with *A Star search* algorithm (as shown in Fig.4-b). Hence, a normal person would like to make its motion be more like to its planned path, and the E_{global} can be computed by:



(a) Online learned scene struture (b) Planned path of target 201

Fig. 4. Global scene structure energy. Given the online learned scene structure (Fig.a), we can find possible planned paths of person 201 with A* algorithm (Fig.b), and these paths can be utilized to compute global scene structure energy in Eq.(2).

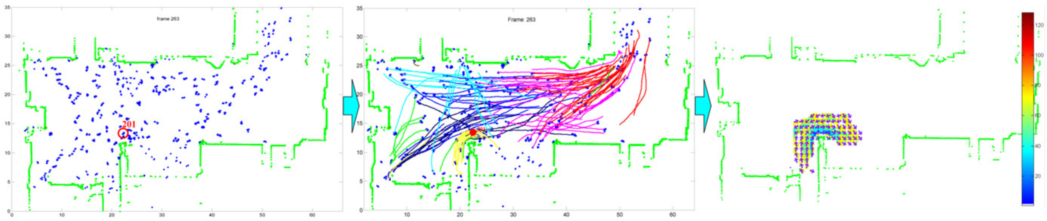
$$\begin{aligned} E_{global}(\mathbf{v}_{i,t}, \mathbf{p}_{i,t}) = \sum_{l=1}^{N} w_l \times \\ \exp(-|| \frac{\mathbf{v}_{i,t}}{||\mathbf{v}_{i,t}||} - \frac{\partial L^l_{i,t}(\mathbf{p}_{i,t})}{\partial x \partial y} ||^2 / 2\sigma_1^2), \end{aligned} \tag{2}$$

where $\frac{\partial L^l_{i,t}(\mathbf{p}_{i,t})}{\partial x \partial y}$ is the tangent vector of $L^l_{i,t}(x,y)$, and it denotes the velocity vector of $L^l_{i,t}(x,y)$ at position $\mathbf{p}_{i,t}$. σ_1 is a constant parameter and w_l is the weight of the possible planned trajectories, which is depend on the similarity between person's current trajectory and this planned one. In this research, we utilize the approach of Wang *et al.* [12] to measure the similarity of two trajectories, please refer it for more details. Once the weights of planned trajectories are very small, we throw them and stop making new path planning for these exits/entrances.

On the other hand, the local crowd flow energy should be also computed. A normal person in a particular crowd flow has to makes its motion be close to the crowd flow. Hence, the motion distribution for a specific crowd flow should be extracted firstly. Given the crowd flow $S_p(t)$, where $\mathbf{x}_{i,t} \in S_p(t)$ (as shown in Fig.5-b), its motion distribution can be computed by:

$$\begin{aligned} \mathfrak{V}_{S_p(t)}(\mathbf{p}_{i,t}) = \exp(- < (\widetilde{v_x}(\mathbf{p}_{i,t}), \widetilde{v_y}(\mathbf{p}_{i,t})), \\ (\cos(\alpha^*_{S_p(t)}(\mathbf{p}_{i,t}), \sin(\alpha^*_{S_p(t)}(\mathbf{p}_{i,t})) > / \eta_v), \end{aligned} \tag{3}$$

here $<,>$ stands for dot product, η_v is a constant parameter, $(\widetilde{v_x}(\mathbf{p}_{i,t}), \widetilde{v_y}(\mathbf{p}_{i,t}))$ is the velocity expectation of cluster



(a) Target 201 in frame 263 (b) Crowd flow (c) Motion distribution

Fig. 5. Crowd flow energy. Given the online learned crowd flow (Fig.b), we can compute its motion distribution (Fig.c). With the help of motion distribution, crowd flow energy can be easily computed by Eq.(5).

$S_p(t)$ at a particular position, and $\alpha^*_{S_p(t)}(\mathbf{p}_{i,t})$ is the principal component in the distribution of flow orientation:

$$\theta_{S_p(t)}(\mathbf{p}_{i,t}) = \sum_{m=1}^{M} \pi_m N(\alpha_{S_p(t)}(\mathbf{p}_{i,t}); \mu, \sigma), \tag{4}$$

where Eq.(4) is GMM model and its parameters π_m can be obtained through EM iteration. An example of Eq.(3) is shown in Fig.5-c, the color denotes the speed, and the arrows display the principal orientation.

Hence, the local crowd flow energy E_{local} can be computed by:

$$E_{local}(\mathbf{v}_{i,t}, \mathbf{p}_{i,t}) = \exp(-||\mathbf{v}_{i,t} - \mathfrak{V}_{S_p(t)}(\mathbf{p}_{i,t})||^2/2\sigma_2^2), \tag{5}$$

where σ_2 is a constant parameter.

C. Centrifugal Force of Pedestrians

For the pedestrian i, the repulsive effects from pedestrian j depend not only on the relative velocity between them, but also on the distance of them (i.e. the headway), and hence these effects can be expressed by a force term with the following form:

$$\mathbf{F}_{ij} = m_i \mathbf{a}_{ij} = -m_i f(\mathbf{v}_{ij}, ||\mathbf{p}_{ij}||)\mathbf{e}_{ij}, \tag{6}$$

where $\mathbf{a}_{ij}$ is the acceleration of pedestrian i caused by pedestrian j, m_i the mass of pedestrian i; $f(\mathbf{v}_{ij}, ||\mathbf{p}_{ij}||)$ is the function of $\mathbf{v}_{ij}$, and $||\mathbf{p}_{ij}||$ to be determined. $\mathbf{p}_{ij}$ is the distance between pedestrian i and j, and it should be:

$$\mathbf{p}_{ij} = \mathbf{p}_j - \mathbf{p}_i. \tag{7}$$

$\mathbf{v}_{ij}$ denotes the projection of the relative velocity of pedestrian i and j in the direction $\mathbf{e}_{ij}$, and can be computed by:

$$\mathbf{v}_{ij} = \frac{1}{2}[(\mathbf{v}_{i,t} - \mathbf{v}_{j,t}) \cdot \mathbf{e}_{ij} + ||(\mathbf{v}_{i,t} - \mathbf{v}_{j,t}) \cdot \mathbf{e}_{ij}||], \tag{8}$$

$$\mathbf{e}_{ij} = \frac{\mathbf{p}_{ij}}{||\mathbf{p}_{ij}||}. \tag{9}$$

From Eq.(7), we can obtain

$$||\mathbf{a}_{ij}|| = f(\mathbf{v}_{ij}, ||\mathbf{p}_{ij}||). \tag{10}$$

According to the proof of [19], there should be

$$\frac{|\mathbf{a}_{ij}|||\mathbf{p}_{ij}||}{\mathbf{v}_{ij}^2} = C, \tag{11}$$

where C is a constant depending on the pedestrian's character. For simplicity, we assume $C = 1$, and obtain

$$\mathbf{F}_{ij} = -m_i \frac{\mathbf{v}_{ij}^2}{||\mathbf{p}_{ij}||}\mathbf{e}_{ij}. \tag{12}$$

If $(\mathbf{v}_{i,t} - \mathbf{v}_{j,t}) \cdot \mathbf{e}_{ij} > 0$, i.e., pedestrian i gets close to pedestrian j, the repulsive effects occur. However, if $(\mathbf{v}_{i,t} - \mathbf{v}_{j,t}) \cdot \mathbf{e}_{ij} < 0$, i.e., pedestrian j walks faster than pedestrian i, there are no repulsive effects. Larger $\mathbf{v}_{ij}$ creates greater repulsive effects in the former case. We assume that pedestrians react to those who are within their angle of view and the field of vision is $180°$, this situation can be characterized by the following coefficient:

$$K_{ij} = \frac{1}{2} \times \frac{\mathbf{v}_{i,t} \cdot \mathbf{e}_{ij} + ||\mathbf{v}_{i,t} \cdot \mathbf{e}_{ij}||}{||\mathbf{v}_{i,t}||}. \tag{13}$$

Hence, the *Centrifugal Force* between pedestrian i and j should be given in the form

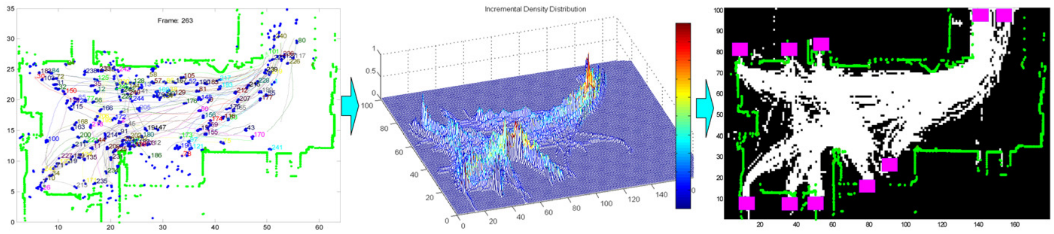
$$\mathbf{F}_{ij} = -m_i K_{ij} \frac{\mathbf{v}_{ij}^2}{||\mathbf{p}_{ij}||}\mathbf{e}_{ij}. \tag{14}$$

Therefore, for pedestrian i, its centrifugal force of J neighboring persons can be computed by:

$$F_{cent}(\mathbf{v}_{i,t}, \mathbf{p}_{i,t}) = || - \sum_{j=1}^{J} m_i K_{ij} \frac{\mathbf{v}_{ij}^2}{||\mathbf{p}_{ij}||}\mathbf{e}_{ij}||. \tag{15}$$

V. Tracking and Learning

In order to compute the abnormal energy, we need to obtain the tracking results of persons, scene structure and crowd flow model. In the proposed system, these information can be computed by "an adaptive tracking-learning loop": we utilize the obtained tracking results to online learn the scene structure and crowd flow model. At the meanwhile, the learned statistical scene model in turn can be used to assist in tracking. Therefore, this mode of co-operation between tracking and learning becomes "an adaptive loop", which can not only dynamically reflects the change of statistical scene model, but also maintain the robust tracking in the extremely crowded environment. Moreover, the entire process is completely online and automatic that require no human intervention. In this section, we will provide the details about them.



(a) Tracking results (b) Density distribution (c) Learned scene structure

Fig. 6. Global scene structure learning. With the help of tracking results (Fig.a), we can compute incremental density distribution (Fig.b), and the global scene structure can be obtained by thresholding the incremental density distribution and the gradient searching.

A. Tracking Model

Consider the state $\mathbf{x}_{i,t} = (\mathbf{p}_{i,t}, \mathbf{v}_{i,t})$ of person i at time t, with its measurement $\mathbf{z}_{i,t}$, where measurement $\mathbf{z}_{i,t}$ is the foreground laser points set after Mean-shift clustering [20], we should estimate its state as

$$\hat{\mathbf{x}}_{i,t} = \arg\max_{\mathbf{x}_{i,t}} p(\mathbf{x}_{i,t}|\mathbf{z}_{i,t}). \tag{16}$$

The posterior probability $p(\mathbf{x}_{i,t}|\mathbf{z}_{i,t})$ can be computed by a Bayesian recursion as

$$p(\mathbf{x}_{i,t}|\mathbf{z}_{i,t}) = \gamma p(\mathbf{z}_{i,t}|\mathbf{x}_{i,t}) \times \int p(\mathbf{x}_{i,t}|\mathbf{x}_{i,t-1}) p(\mathbf{x}_{i,t-1}|\mathbf{z}_{i,t-1}) d\mathbf{x}_{t-1}, \tag{17}$$

where γ is the normalization constant, $p(\mathbf{z}_{i,t}|\mathbf{x}_{i,t})$ the similarity between target's state and measurement (observation model), and $p(\mathbf{x}_{i,t}|\mathbf{x}_{i,t-1})$ is the transition probability (dynamic model).

Based on our previous work [21], [22], we utilize particle filter technique to compute equation (16) and (17). For more details about dynamic model $p(\mathbf{x}_{i,t}|\mathbf{x}_{i,t-1})$ and observation model $p(\mathbf{z}_{i,t}|\mathbf{x}_{i,t})$, please see the above two references. In addition, we utilize the online learned statistical scene model to improve the tracking results in further: the crowd flow is used for controlling the dynamic model, and the scene structure is for dealing with the problem of uncertain measurements. For more details about this part, please refer [21]. After re-weighting and re-sampling in the particle filter process, it is easy for us to obtain the position $\mathbf{p}_{i,t}$ and velocity $\mathbf{v}_{i,t}$ of each pedestrian.

B. Global Scene Structure Learning

With the proceeding of tracking, it is easy for us to obtain a large number of trajectories of pedestrians, and we can utilize them to learn the global scene structure. Firstly, the spatial extent of these trajectories should be learned and it can be described by the density distribution. Given all the trajectories $L_{i,t}(x, y)$ we obtained at time t, the density distribution at position (x, y) can be estimated as:

$$\mathfrak{D}_{global}(x, y, t) = \sum_{(x_i, y_i) \in L_{i,t}} \sum_{L_{i,t} \in \Omega} \exp(-\|(x - x_i, y - y_i)\|^2/\eta_d), \tag{18}$$

where Ω is all the trajectories we have obtained at time t, η_d a constant parameter. Therefore, the dominant paths of the scene were easily extracted by thresholding the incremental global density distribution (as shown in Fig.6).

On the other hand, the exit and entrance of the scene are two important scene properties, which are also called sources/sinks. The sinks/sources can be easily detected from the global density distribution $\mathfrak{D}_{global}$. As shown in Fig.6-c, the sinks/sources usually occur at the region of great change of the global density distribution after thresholding. In addition, the changed direction must follow the principal orientation of the crowd flow. Hence, the location $\mathbf{p}^*_{sink}$ of sinks/sources can be easily detected by a gradient searching at the principal orientation of density distribution:

$$\mathbf{p}^*_{sink} = \arg\max_{\mathbf{p}_{sink}} (< \nabla \mathfrak{D}_{S_p(t)}(\mathbf{p}_{sink}, t), \vec{v}_p/|\vec{v}_p| >), \tag{19}$$

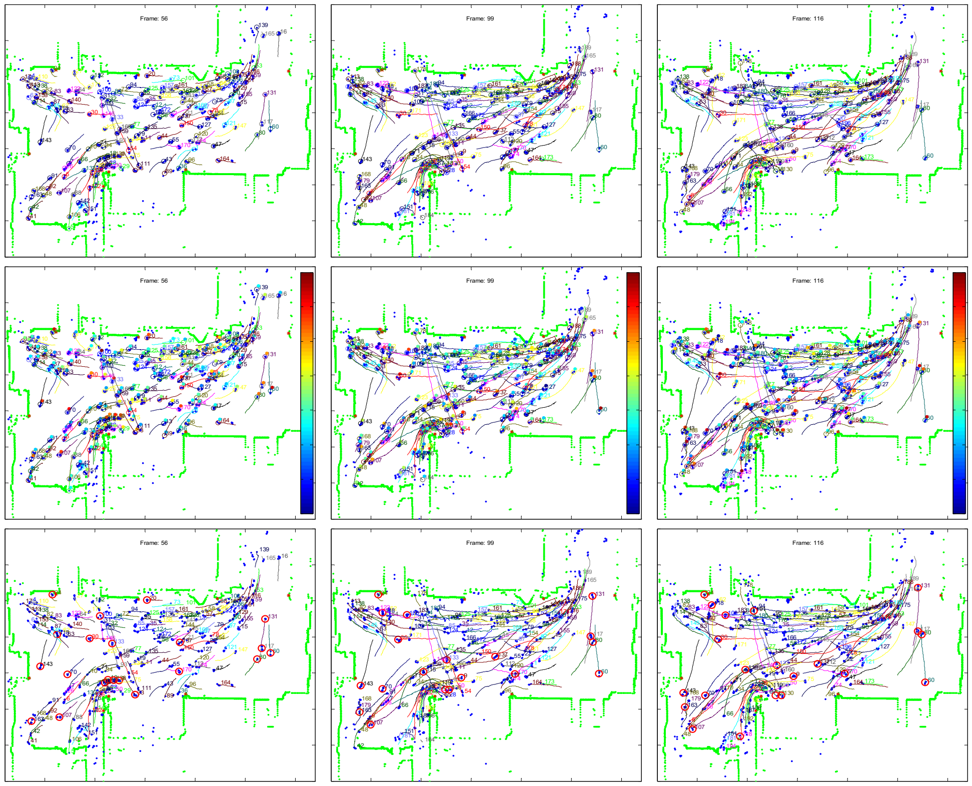
where where $\vec{v}_p$ is the principal direction of the crowd flow $S_p(t)$, and $\mathfrak{D}_{S_p(t)}$ is its density distribution.

An example is illustrated in Fig.6, as the 1090 frames proceeded, the dominant paths and sinks/sources of the scene were obtained, and this scene structure map can be utilized to assist in tracking and measure the pedestrians' abnormal energy.

C. Crowd Flow Learning

The crowd flow $S_p(t)$ can be seen as a group of persons who have similar activities and spatial information, they are usually going to the same destination with similar velocity. Hence, once we obtain the trajectories of pedestrians at time t, they are should be clustered into n clusters $\{S_p(t)\}_{p=1}^n$. In order to dynamically reflect the change of crowd flow, the clustering must be online and can be seen as a function of time t. We consider each cluster $S_j(t)$ as a moving hyperplane. Thus, we can model a union of n hyperplane in $\mathbb{R}^D$, where $S_p(t) = \{\mathbf{x} \in \mathbb{R}^D : \mathbf{b}_p^\top(t)\mathbf{x} = 0\}, p = 1, ..., n$, where $\mathbf{x}$ is the person state, $\mathbf{b}(t) \in \mathbb{R}^D$, as the zero set of a polynomial with time varying coefficients using normalized gradient descent. Then the hyperplane normals are estimated from the derivatives of the new polynomial at each trajectory. Lastly, the trajectories are grouped by clustering their associated normal vectors. For more details about this part, please refer [23], [21].

On the other hand, we have to control the influence of the crowd flow in the energy function. Let us come back to



Frame 56 Frame 99 Frame 116

Fig. 7. The results of the proposed system. The first row is the tracking results, and the second is the abnormal energy, where the color shows abnormal energy value, the darker the greater value (as shown in the color bar). The abnormality detection results are shown in the last row. Some suspicious persons could be easily detected by our system, please note that person 131, 117, 60 and 80 in frame 56, person 24 in frame 99 and person 9 in frame 116. For more results of our system, please see our supplementary video.

Eq.(1), for the pedestrian i, the influence from crowd flow $S_p(t)$ will be depend on the density distribution of this crowd flow, and it can be computed by:

$$\alpha_{i,t} = 1 - \exp(-\mathfrak{D}_{S_p(t)}(\mathbf{p}_{i,t}, t)), \tag{20}$$

where pedestrian $\mathbf{x}_{i,t} \in S_p(t)$, $\alpha_{i,t} \in (0,1)$, and this equation can be understood like this: at a specific time t, a person $\mathbf{x}_{i,t}$ in crowd flow $S_p(t)$ were walking in position $\mathbf{p}_{i,t}$. If this area in this crowd flow is quite crowded, this person's activities will be greatly influenced by this crowd flow. In contrast, this influence will be little.

In summary, we can utilized the online learned scene structure and crowd flow to compute Eq.(2) and Eq.(5). Then, the *abnormal energy* $E(\mathbf{v}_{i,t}, \mathbf{p}_{i,t})$ for pedestrian i can be computed by Eq.(1). Lastly, the abnormal activities can be detected by thresholding the *abnormal energy* of each person.

VI. Experiments and Results

We applied the proposed system to the real scene: lobby of JR subway station (about 60m×35m). Eight single-row laser scanners (LMS291) produced by SICK was utilized. They were set above 10cm on the ground surface and performing the horizontal scanning with a frequency of 37 fps. The selected data used for evaluation was from 7:00am to 8:30am when was a very busy time in Tokyo. We selected some representative sequences, and labeled some obvious abnormal activities in them. Hence, the $\alpha_{i,t}$ in Eq.(1) and the threshold of abnormal energy could be easily learned by the regression analysis. In addition, σ_1 and σ_2 in Eq.(2) and Eq.(5) was set to 0.3. In this section, we will present our experimental results and perform the quantitative evaluation.

A. Results

Some selected results of our system are shown in Fig.7. The first row is the tracking results, and the second is the abnormal energy, where the color shows abnormal energy value, the darker the greater value (as shown in the color bar). The abnormality detection results are shown in the last row. From this figure, we can see that some suspicious persons could be easily detected by our system, such as person 131, 117, 60 and 80 in frame 56 (walking on the closed road), person 24 in frame 99 (what was it doing?), person 9 in frame 116 (it was not following other persons and walking in a strange path) and etc. For more results of our system, please see our supplementary video.

B. Quantitative Evaluation and Comparison

In order to perform quantitative evaluation, we invite three persons to label the abnormal activities in our data. One of

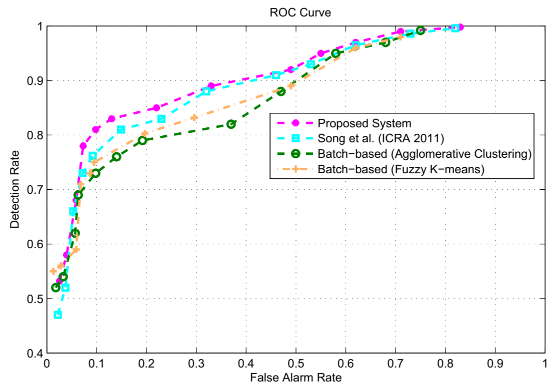


Fig. 8. Quantitative comparison among four methods. This figure shows the ROC curves of four methods. From this figure, we can see that the proposed system has the better performance than the other three methods.

them has the academic background, and the other two do not have any academic background of this area. We tested our system with 3000 frames, and the ROC curve of the proposed system is shown in Fig.8. In addition, a quantitative comparison was also conducted among four methods: Fuzzy K-means, Agglomerative Clustering, Song *et al.* [14] and ours. For the first two competing algorithms, the abnormal activities were detected by finding the outlier trajectories of the clusters. The details of this comparison are also shown in Fig.8. From this figure, we can see that the proposed system has the better performance than the other three methods.

VII. Conclusion

In this paper, we propose a novel abnormality detection model and develop an fully online surveillance system, which can a cover large area, perform robust tracking, semantic scene learning and online abnormal detection. The proposed model incorporates the intensively explored semantic scene knowledge and social interactions among persons, and the experimental results have demonstrated its feasibility and robustness. Actually, for the laser scanner data, it is difficult to give a clear and unify definition about abnormal activity, therefore, how to scientifically evaluate the detected results will become an very important issue and it will become the focus point of our future research.

VIII. Acknowledgments

This work was supported by a Grant-in-Aid for Young Scientists (23700192) and "Strategic Project to Support the Formation of Research Bases at Private Universities": Matching Fund Subsidy by Japan's Ministry of Education, Culture, Sports, Science, and Technology (MEST). This work was also partially funded by East Japan Railway Company and Microsoft Research.

References

[1] H. Zhao and R.Shibasaki, "A novel system for tracking pedestrians using multiple single-row laser range scanners," *IEEE Transactions on Systems, Man and Cybernetics, part A*, pp. 283–291, 2005.

[2] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 15, pp. 1–72, 2009.

[3] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Abnomaly detection in crowded scenes," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1975–1981, 2010.

[4] S. Wu, B. E. Moore, and M. Shah, "Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 2054–2060, 2010.

[5] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 555–560, 2008.

[6] R. Mehran, A. Oyama, and M. Shah., "Abnormal crowd behavior detection using social force model," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 935–942, 2009.

[7] L. Patino, F. Bremond, M. Evans, A. Shahrokni, and J. Ferryman, "Video activity extraction and reporting with incremental unsupervised learning," *Proc. IEEE International Conference on Advanced Video and Signal-Based Surveillance*, pp. 54–61, 2010.

[8] X. Wang, K. Ma, G. Ng, and E. Grimson, "Trajectory analysis and semantic region modeling using a nonparametric bayesian model," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[9] W. Hu, D. Xie, Z. Fu, W. Zeng, and S. Maybank, "Semantic-based surveillance video retrieval," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, no. 4, pp. 1168–1181, 2007.

[10] G. Pusiol, F. Bremond, and M. Thonnat, "Trajectory based activity discovery," *Proc. IEEE International Conference on Advanced Video and Signal-Based Surveillance*, pp. 75–82, 2010.

[11] Z. Fu, W. Hu, and T. Tan, "Similarity based vehicle trajectory clustering and anomaly detection," *Proc. IEEE International Conference on Image Processing*, pp. 1133–1136, 2005.

[12] X. Wang, K. Tieu, and E. Grimson, "Learning semantic scene models by trajectory analysis," *Proc. European Conference on Computer Vision*, pp. 110–123, 2006.

[13] A. Basharat, A. Gritai, and M. Shah, "Learning object motion patterns for anomaly detection and improved object detection," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[14] X. Song, X. Shao, R. Shibasaki, H. Zhao, J. Cui, and H. Zha, "A novel laser-based system: Fully online detection of abnormal activity via an unsupervised method," *Proc. IEEE International Conference on Robotics and Automation*, pp. 1317–1322, 2011.

[15] B. Zhao, F. Li, and E. P. Xing, "Online detection of unusual events in videos via dynamic sparse coding," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 3313–3320, 2011.

[16] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 3449–3456, 2011.

[17] J. Li, S. Gong, and T. Xiang, "Scene segmentation for behaviour correlation," *Proc. European Conference on Computer Vision*, pp. 383–395, 2008.

[18] R. Mehran, B. E. Moore, and M. Shah, "A streakline representation of flow in crowded scenes," *Proc. European Conference on Computer Vision*, pp. 439–452, 2010.

[19] W. Yu, R. Chen, L. Dong, and S. Dai, "Centrifugal force model for pedestrian dynamics," *Physical Review E*, vol. 72, no. 2, p. 112C119, 2005.

[20] D. Comaniciu and P. Meer, "Distribution free decomposition of multivariate data," *Pattern Analysis and Applications*, vol. 38, pp. 22–30, 1999.

[21] X. Song, X. Shao, H. Zhao, J. Cui, R. Shibasaki, and H. Zha, "An online approach: Learning-semantic-scene-by-tracking and tracking-by-learning-semantic-scene," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 739–746, 2010.

[22] X. Song, J. Cui, X. Wang, H. Zhao, and H. Zha, "Tracking interacting targets with laser scanner via on-line supervised learning," *Proc. IEEE International Conference on Robotics and Automation*, pp. 2271–2276, 2008.

[23] R. Vidal, "Online clustering of moving hyperplanes," *Proc. Neural Information Processing Systems*, pp. 1433–1440, 2006.