1722　レーザレンジセンサを用いた駅内の旅客の追跡と流動の計測

邵 肖偉　（東京大）　　正［土］　○柴崎 亮介　（東京大）

尾住 秀樹　（東日本旅客鉄道㈱ JR東日本研究開発センター）

坂本 圭司　（東日本旅客鉄道㈱ JR東日本研究開発センター）

People Surveillance and Tracking via Range Sensor at Railway Stations

Xiaowei SHAO, Tokyo Univ. 4-6-1 Komaba, Meguro-ku, Tokyo
Ryosuke SHIBASAKI, Tokyo Univ.
Hideki OZUMI, East Japan Railway Co.
Kiyoshi SAKAMOTO, East Japan Railway Co.

In this study we present how to exploit laser range sensors for 2D/3D crowd surveillance and tracking applications. Multiple single-row laser scanners are utilized to track the movement of pedestrians in a crowded scene, while swinging laser scanners and 3D range sensor provide a 3D representation of crowd. Quantified crowdness analysis is conducted from various aspects to indicate the situation inside the surveillance area according to the localization/tracking results of passengers. Several experiments were conducted at subway stations in Tokyo and experimental results demonstrate the effectiveness of our system.

Keywords : range sensor, people surveillance, crowd tracking, railway station

1. Introduction

In recent years, due to safety and security concerns the demand for sensing and analyzing people flow inside a railway station is significantly growing in many important applications, such as building security, space utilization, disaster management and crowd flow control.

Traditional video camera based surveillance systems have been extensively studied and widely used in the past several years. Most of such systems are built on video-based object detection and tracking techniques and can be quite effective for tracking limited objects. However, when the number of object inside the surveillance area increases, usually the performance of such systems may decrease because the distance among people becomes small and it becomes difficult to separate them from each other due to occlusion. In addition, it is insufficient to derive the location of objects in real dimension, so the cooperation among multiple video cameras is also a complicated task. Therefore, crowd surveillance remains a challenging issue for many video camera based systems, especially for congested scenes such as a central railway station in rush hour.

In our proposed system, range scanners are utilized for data collection instead of video cameras. Being a kind of measurement instrument which has been fully developed in recent years, the laser range scanner has received increasing attention for solving surveillance and tracking problems [1]-[4].

In general, range sensor works by actively emitting invisible light beams at controlled directions and measures distances of nearby objects by computing their time of flight. Compared with traditional video cameras, it cannot provide detailed color information of objects, but it has special advantages in positioning because it measures distance directly and records the position of objects as the actual distance. In addition, the range sensor is independent of many factors such as illumination, object color and texture, which ensures the robustness of our proposed system.

To provide a better spatial awareness of people inside the scene, in this study we address the issue of crowd surveillance and tracking at railway stations by utilizing range sensors in three different ways. Multiple single-row laser scanners are utilized to track the movement of pedestrians in a congested scene, while swinging laser scanners and Kinect range sensor could provide a 3D representation of crowd. Quantified crowdness analysis is conducted from different aspects to indicate the situation inside the surveillance area according to the localization/tracking results of passengers. Several experiments were conducted at subway stations in Tokyo and experimental results demonstrate the effectiveness of our system.

2. Measurement System

In this section, we provide a briefly introduction of our sensing systems and their characteristics. In all three systems, multiple sensors are utilized to improve the sensing ability of our systems in different aspects.

Horizontal scanning system. Single-row laser range scanners (LMS 200) produced by IBEO Company is exploited. During one scan, the scanner profiles 361 range

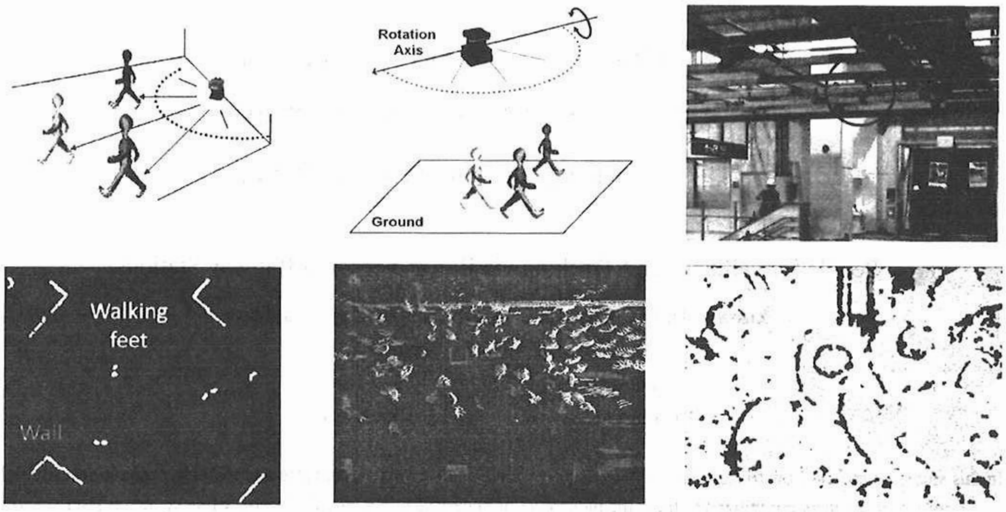


Fig. 1. Measurement system. Left: horizontal scanning system and sample data; middle: swinging scanning system and sample data at a platform; right: Kinect and a sample range image near a stair

distances equally in 180 degrees on its scanning plane. The collected range data can be converted into 2D points in Cartesian coordinates. The laser scanner has a maximum range distance of 80 m and an average distance error of 4 cm. Its scanning frequency is 37.5 Hz.

This system utilizes multiple laser scanners so it can cover a large area such as an experimental site of 60m x 20m. Each laser scanner is located at a separate position and is controlled by a client computer. All client computers are connected to a server computer, which gathers the laser points of moving objects from all clients and tracks trajectories. The integrated frame rate in the server is 20 FPS, sufficient for most surveillance and tracking applications. During experiments, the laser scanners were set making horizontal scans at 10 cm above ground level, targeting the ankle of a typical walking person.

Swinging sensing system. By installing a small and light single-row laser range scanner (Hokuyo UTM-30LX) on a programmable swinging platform, this sensing system is capable of collecting 3D range data of objects inside the surveillance area, at a speed of 0.7 – 1.0 second per swing. UTM-30LX scans objects within 30 m with an average distance error of 3 cm, and a scanning frequency of 40 Hz. In addition, it has a better spatial resolution, profiling 1080 range distances equally in 270 degrees. Multiple scanners are exploited to increase the temporal resolution and spatial coverage of this sensing system.

Kinect range sensor. Kinect is a new type of sensing input device developed by Microsoft. It embeds a 3D range sensor and can collect high resolution 640 x 480 range images at 30 FPS with a low cost. Kinect scans objects from about 0.7– 6 m and its precision decreases when the distance increases. In addition, its field of view is somewhat limited (horizontal: 57 degree; vertical: 43 degree).

3. Crowd Surveillance and Tracking

According to the characteristics of these sensing systems, different approaches are proposed for crowd surveillance and tracking applications.

3.1 Horizontal scanning system

Horizontal scanning system collects 2D points inside the scene and our task is to track the feet of each pedestrian. We notice that a specific pattern is generated by walking feet, and this pattern remains explicit and effective even in dense clutter. Based on this feature, a walking model is constructed and then applied in our tracking algorithm. Particle filter is exploited to track the status of each pedestrian.

During tracking, the number of people can be time-varying, and the total number is quite large sometimes. To deal with a large number of interactive targets simultaneously while keeping the computation complexity at an acceptable level, we propose crowd tracking via interactive multiple particle filters. In addition, spatio-temporal correlation analysis based on cluster information is exploited to handle a variable number of pedestrians. Post processing that includes trajectory connection and extended Kalman smoothing based trajectory completion is optionally available in our system.

To examine the performance of our horizontal scanning system, a field experiment at a railway station in Tokyo was conducted, covering an area of about 60 m x 20 m using 8 laser scanners. Here two cases were studied. The first one started at 7:00 am and lasted for 5 minutes. There were 387 pedestrians in total, of which 367 (94.8%) were successfully tracked. The trajectories of all pedestrians are shown in Fig. 2. The second one started at 8:00 am and lasted for 10 minutes. According to our statistics, there were 2608 pedestrians in total, and the number of targets reached as high as 170 at peak time. The tracking accuracy is summarized as follows. For all 2608 pedestrians, 2335 (89.5%) of them were successfully tracked. For the 273 tracking failures, 64 of them involved breaks, 180 involved hijacking, and 29 were for other reasons. A sample image of tracking results is shown in Fig. 2.

The horizontal scanning system is effective for tracking moving person. However, when a group of people stand still we found the occlusion problem becomes so severe and the people inside the group can be barely seen. Therefore, we tried to extend the traditional 2D laser data into a 3D way to reduce the extent of occlusion, at the cost of decrease in temporal resolution. That idea leads to our swinging sensing system.

3.2 Swinging sensing system

In the swinging sensing system, 3D points are captured from the swinging sensors. For each swing, we convert the point data into a range image. A histogram based method is proposed to extract background information from the original data, by finding the most frequent distance for each pixel in the range image. Since multiple sensors are employed, it is necessary to perform calibration to determine the pose of each sensor so that the point data can be integrated. We complete this task via a two-step semi-auto calibration scheme. The first step match the reference ground plane with manually selected area and the second one perform a common 2D calibration.

After finishing pre-processing steps, we now have the 3D points from multiple laser scanners. These points can be integrated directly since they represent the position of objects in real-world scenes. To extract people in the crowd, an improved mean-shift algorithm was developed. Traditional mean-shift algorithm is computational intensive due to the large number of query of nearby 3D points. We refine the clustering algorithm correspondingly from two aspects: 1) rasterize 3D points and use neighbor grids of candidates during iteration to accelerate the computation; 2) cluster the 3D points by using their position in x-y plane while the height is considered as the weight correspondingly. In this way, our proposed method is more efficient and the clustering accuracy is also well preserved.

Based on the extraction result of passengers, we conducted crowd analysis from three aspects: 1) number of people; 2) density of people; 3) personal space of people. First some regions of interests (ROIs) are specified manually and the numbers of people and the density inside ROIS are calculated. Meanwhile, for each person we compute the average distance from current object to his/her nearest k neighbors and define it as personal space. The histogram of personal space is calculated and fitted by a g-distribution, in order to provide a quantized description about the crowdness in the surveillance area.

An experiment was conducted in 2010 at a central railway platform of JR station in Tokyo. Swinging laser scanners were mounted near the cell at a height of about 5.30 m. The rotation angle of the swinging platform ranges from -55 deg to 55 deg, at a speed of about 1.0 second per swing. Two data sets were utilized to verify the performance of our proposed system. One was captured from 8:30 to 8:40 in the morning while the other lasted from 18:30 to 18:35. Fig. 3 shows an example of people extraction result in ROIs and the corresponding histogram of personal space. It can be seen that in rush hour for most people their personal space is less than 1.0 m. We accumulate the histogram of personal space and its fitting result is also given in Fig. 3.Our analysis shows that 8:30 case it has a lower average personal space and sharper peak than 18:30 case, which indicate it is more congested in the former case. Fig. 4 shows the number and density distribution of people in ROIs in 10 minutes. There are total 4 ROIs, where #1 and #2 are waiting areas, and #3 and #4 are regions close to stairs and escalators. It can be seen from Fig. 4(a) that ROI#1 and #2 have sharp peaks, which indicate people enter and leave this area very fast after the arrival of a train. ROI #3 and #4 have slow but much longer peaks which means people get congested in these ROIs. Fig. 4(b) shows the density distribution of ROIs. We divided the density into 6 levels and the percentage of different levels is displayed. It is clear that ROI #3 and #4 (especially #3) are more crowded than ROI #1 and #2.

3.3 Kinect range sensor

Compared with swinging sensing system, the advantage of Kinect sensor is its high scanning speed (30FPS) and high spatial resolution. By utilizing Kinect sensors, it is possible to conduct 3D crowd tracking.

The range image collected by Kinect is similar to the range data from our swinging sensing system. So similar pre-processing steps can be applied to perform calibration and extracted foreground 3D points. For extracted foreground 3D points, crowd tracking is conducted. We apply our improved mean-shift clustering technique to extract the location of people and try to track each person by combining spatial-temporal consistency and their neighborhood.

A field experiment was conducted in 2012 to evaluate the performance of Kinect sensing system, where more than 30 people anticipated this experiment. Two test cases were analyzed and the overall tracking accuracy of our algorithm is 93.7%. Example of 3D people extraction and tracking are given in Fig. 5.

Kinect has great advantages in acquiring high-quality range images. Meanwhile, Kinect also has some limitations. First, its field of view and sensing range is somewhat limited, which means it can only cover a small area around several square meters. Second, the sensing precision will decrease when the distance increases. In addition, Kinect can only be used for indoor applications. If exposed under sunlight, usually it cannot make a valid estimation about distance because the emitted light beam lies in the spectrum of sunlight but has a much lower strength.

4. Conclusions

In this study, we investigated different ways of sensing and tracking crowd by range sensors, as well as their characteristics and the methodology of data processing. Various experiments show the validity of our proposed system. It can be seen that range sensor has a good performance for crowd surveillance and tracking applications and is able to provide accurate and quantified information of objects of interests. Meanwhile, the ability of existing sensing systems is still limited, so in the future we focus on integration and fusion of different sensing system for a better spatial awareness of people inside the surveillance area.

参考文献

1) Fod, A., et al. : Laser-based people tracking, in IEEE Conference on Robot. Autom., pp. 3024-3029, 2002.

2) Fayad, F. and Cherfaoui, V. : Tracking objects using a laser scanner in driving situation based on modeling target shape, in IEEE Intelligent Vehicles Symposium, pp. 44-49, 2007.

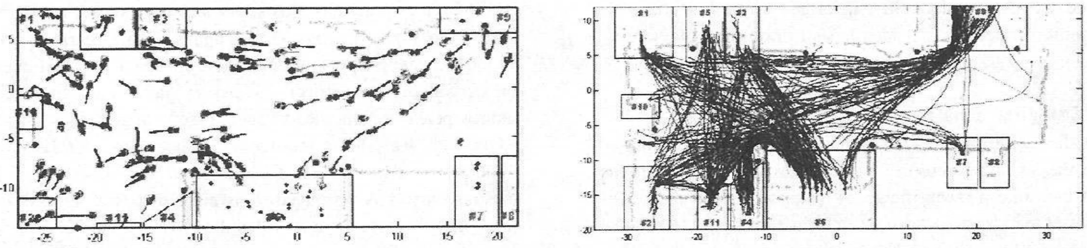


Fig. 2. Crowd tracking via horizontal scanning system. Left: a sample image of tracking result of 8:00 am case; right: trajectories of all pedestrians of 7:00 am case

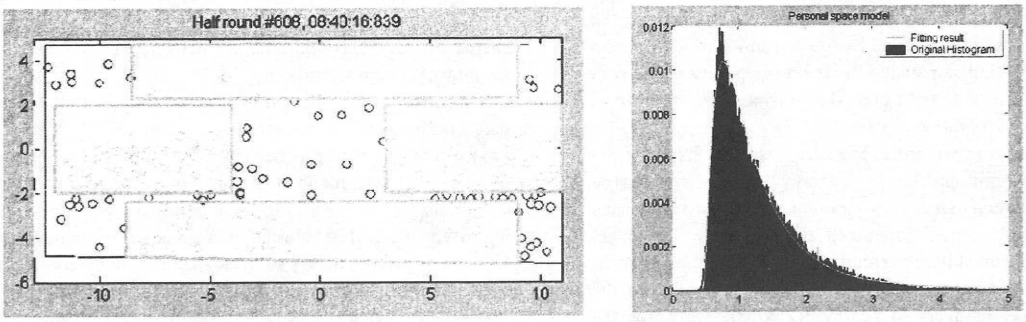
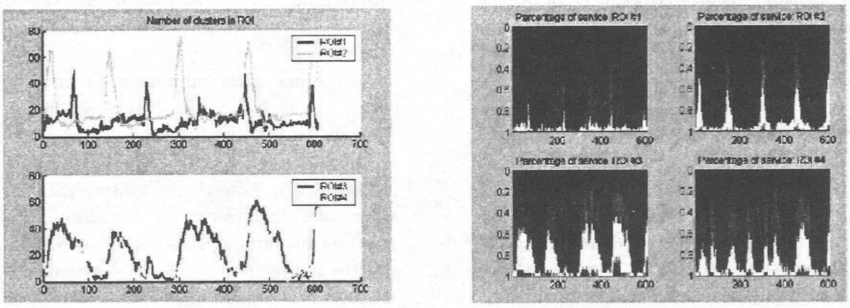


Fig. 3. Crowd surveillance via swinging scanning system. Left: an example of people extraction result in regions of interests; right: histogram of personal space between 8:30~8:40 and corresponding fitting result



(a) Number of people in ROIs

(b) Density distribution in ROIs

Fig. 4. Number of people and density distribution in ROIs

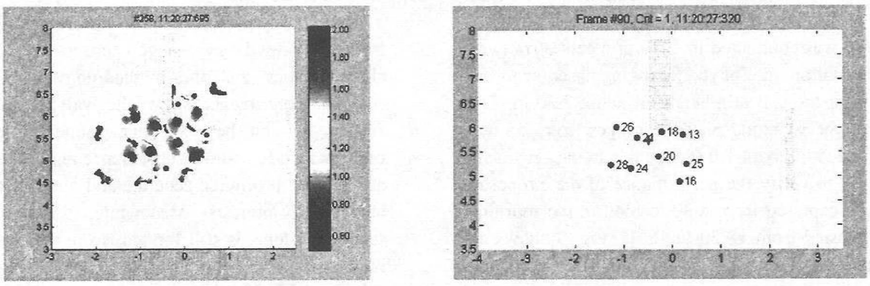


Fig. 5. Crowd tracking via Kinect sensing system. Left: example of people extraction result; right: an example of tracking result in surveillance area

3) Zhao, H., et al. : A laser-scanner-based approach toward driving safety and traffic data collection, IEEE Trans. Intelligent Transportation Systems, vol. 10, no. 3, pp. 534-546,2009.

4) Shao, X., et al. : Tracking a variable number of pedestrians in crowded scenes by using laser range scanners, in IEEE Conference on Systems, Man, and Cybernetics, pp. 1545-1551, 2008.