Sensing an Intersection Using a Network of Laser Scanners and Video Cameras

Huijing Zhao, Jinshi Cui, and Hongbin Zha
Key Lab of Machine Perception (MOE) Peking University, China zhaohj@cis.pku.edu.cn

Kyoichiro Katabira, Xiaowei Shao, and Ryosuke Shibasaki
Control Center for Spatial Information Science University of Tokyo, Japan katabira@csis.u-tokyo.ac.jp

Abstract — In this research, a novel system for monitoring an intersection using a network of single-row laser range scanners (subsequently abbreviated as "laser scanner") and video cameras is proposed. Laser scanners are set on the road side to profile an intersection horizontally from different viewpoints. The contour points of moving objects are captured at a certain horizontal plane with a high scanning rate (e.g., 37 Hz). A laser-based processing algorithm is developed, thus the moving objects entered the intersection are detected and tracked to estimate their state parameters, such as: location, speed, and direction at each time instance. In addition, laser data and processing results are forwarded to an associated video camera, so that a visualization as well as fusion-based processing can be achieved. An experiment in central Beijing is presented, demonstrating that a large quantity of physical dimension and detailed traffic data can be obtained through such a system.

I. Introduction

Analyzing or monitoring traffic behavior at an intersection, such as collecting the traffic speed data, motion trajectory and counts for different kinds of traffic objects (i.e., car, bicycle and pedestrian), is very important. Many researchers in the field of ITS and computer vision have devoted to the development of vision-based systems to collect such data (e.g., [1,2,3,8]). However, vision-based systems suffer from two problems: occlusions and sudden changes in

Digital Object Identifier 10.1109/MITS.2009.933864

illumination. In most of the existing systems, video cameras are required to be set in a restricted position (e.g., in a high position) so that the objects on the road can be monitored with less occlusion. However, such conditions can not be easily achieved at many intersections. If additional construction is required, the setting cost might be too high for a non-regular system. Thus, a new method with easy setting conditions is needed for collecting detailed traffic data at an intersection.

In addition to video camera, laser scanner (also called Lidar, laser range scanner, etc.) is alternative that can detect moving objects and measure their motion trajectories within a horizontal area. Up to now, laser scanners have been broadly used to assist with driving safety (e.g., a laser scanner is set at the front of a car or a robot where it detects obstacles, moving cars, and pedestrians to prevent collisions (e.g., [4,5,10,11,12]). Regarding the constant monitoring of a certain intersection and collecting traffic data in different aspects, the ITS group at the Univ. of Minnesota developed a test bed system that placed a network of radars and laser sensors [9] near a rural intersection. While researches are still required on developing data processing algorithms so as to obtain a comprehensive perception to the dynamic scene.

In this research, we propose a system for sensing an intersection using a network of laser scanners and video cameras. The expected output of the system is the motion trajectories of each moving object that entered the intersection and an estimation to its class (e.g., car, bus, bicycle or pedestrian). This is an extension of our previous work. In [6,13] we proposed a system that using a network of laser scanners to extract the motion trajectories of pedestrians and reported its application of collecting the detailed traffic data at railway stations. In [14] we considered a more challenging situation, an intersection, where different kinds of moving objects and motion patterns exist. An algorithm was proposed of jointly tracking and classifying the moving objects at an intersection using a single laser scanner. However, when facing a large and crowded intersection, a network sensing from different viewpoints is required to reduce occlusions, and algorithms to integrate the sensing data, conduct moving objects detection, tracking and classification are to be developed.

This paper is organized as follows. In section 2, we give an outline of our sensor and data measurement system. In section 3, we address the data analysis for detecting, tracking and classifying moving objects. The experiment is described in section 4, followed by conclusions and future topics in section 5.

II. Outline of the System

A diagram of the system setting is given in Figure 1. In this research, we use the single-row laser scanners, LMS291 by SICK. When setting the scanning angle to 180 degrees and scanning resolution to 0.5 degrees, the sensor has a scan rate of about 37.5 Hz. On each scan, it measures 361 range values at an equal angular interval of 0.5 degrees. The maximal range distance is up to 45 meters (according to our experience), with an average range error of 3cm. According to the beam angle, each range value can be easily converted to a 2D coordinate (called a "laser point") with respect to the sensor's local coordinate system.

Laser scanners are set on roadsides and profiles the intersection at a horizontal plane, e.g. 40 cm above the ground, so that the data of vehicle body and human leg are measured. A number of laser scanners are exploited, so a relatively large environment could be covered while reducing occlusion. Each laser scanner is located at a different position and controlled by a client computer. The client computer collect raw measurements from the laser scanner, perform some preliminary processing on the local scan data, such as background image generation and mobile data extraction.

A background image that contained only motionless objects was generated at each laser scanner. For each sampling angle of range scanning, a histogram was generated using the range values from all range frames being examined. A pick value above a certain distinctness (percentage) was found, which indicates that an object was continuously measured in that direction, so that sounds a still

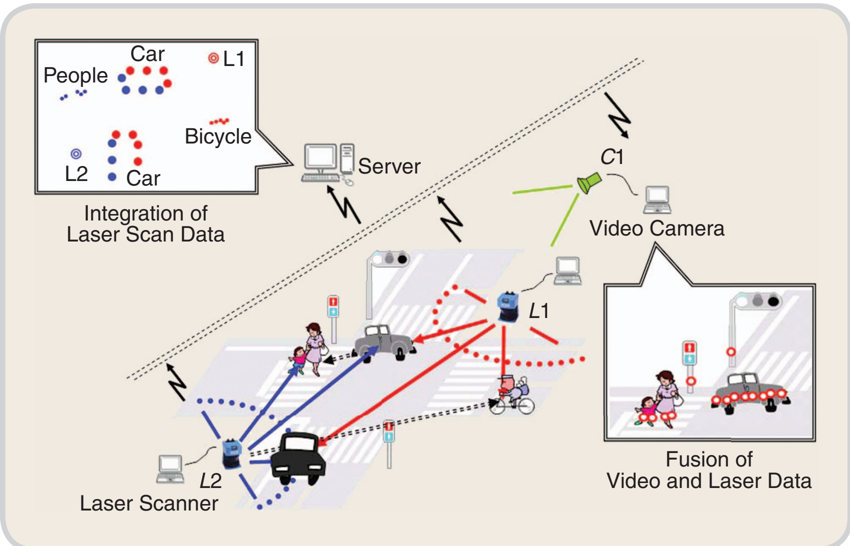

FIG 1 A diagram of the system.

one. The background image was made up of the pick values at all sampling angles. Whenever a new range frame was recorded, background image subtraction was conducted at the level of each range distance. If the difference between the two range values was larger than a given threshold (considering the fluctuation of range measurement), the newly measured range value was extracted as the moving object data.

According to the measurements of the same static objects, a transformation within the horizontal plane (i.e., two translation and one rotation parameter) between the coordinate systems of two neighboring laser scanners can be estimated. Using one laser scanner as the reference frame, all laser points from different laser scanners can be transformed into a common coordinate system. In addition, all client computers are connected through a network to a server computer. The server computer broadcasts its time periodically so as to synchronize the time of all client computers. Examples of data before and after synchronization are shown in Figure 3.

The server computer collects laser scans, as well as local processing results, from all client computers. A processing rate is set at the server computer (20Hz in this research). An integrated frame is generated at each processing interval by collecting the scan of the nearest time stamp from each laser client and converting the coordinates of the laser points to the reference frame. Thus, an integrated frame is an instantaneous cross section of the whole intersection, the laser points of which depict the visible contour of both static and mobile objects in the scanning plane. The detection and tracking of moving objects is conducted by the server computer by using the integrated frames of laser data.

On the other hand, video cameras can also be associated. The data from video camera are not restricted to any horizontal plane, so that 3D information of the extracted moving objects can be estimated. A fusion of video with laser can compensate each other, and achieve an advanced perception to the environment.

Figure 2 describes a design of the system architecture and major data flow. In order to distribute computation cost and reduce network data transfer, for each video camera, server computer extracts corresponding laser points from the integrated frames and forwards them to the client computer. Fusion based data processing is conducted in each video client.

Except calibration and synchronization, which are mainly conducted by server computer, there are three major modules: laser-based moving object detection/tracking, fusion-based moving object detection/tracking and, traffic data process and visualization. Currently, we focus mainly on the development of a laser-based approach, and using video for validation. Other modules will be addressed in future work.

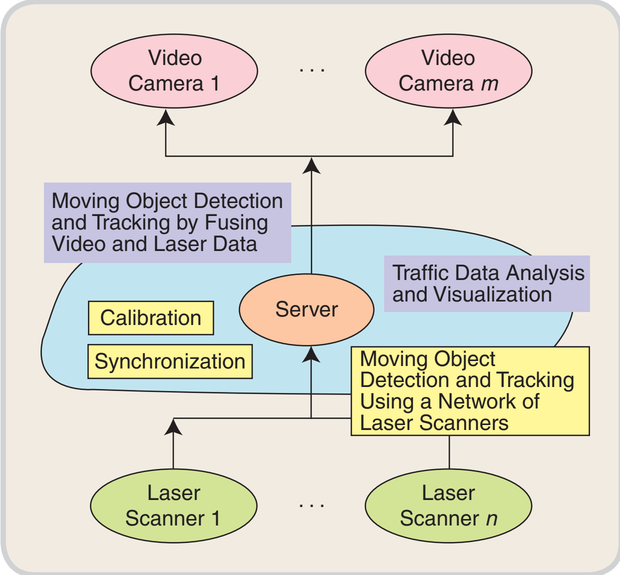

FIG 2 System architecture and major data flow.

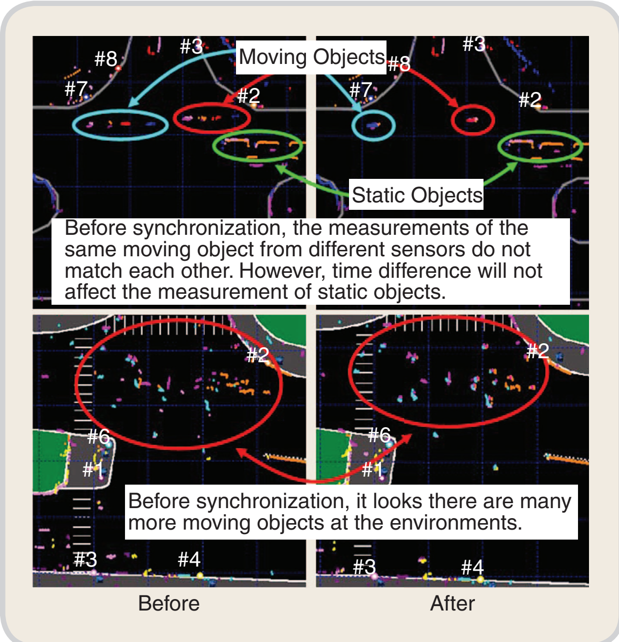

FIG 3 Synchronization of laser scanners.

III. Moving Object Detection and Tracking

A. A Description to the Data

Two pieces of laser data samples are shown in Figures 4 and 5, where the gray lines, representing the road boundaries and safety zones of the intersection, are included for a better understanding of the environment. Three laser scanners were placed on the road side of a three-way intersection, the dimension of which is demonstrated by blue mesh lines with a 10 m step.

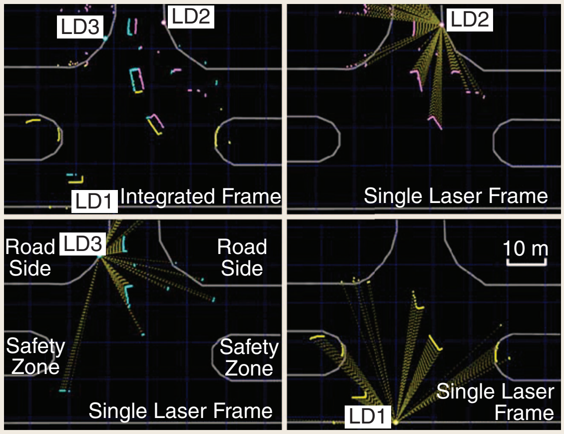

FIG 4 A laser data sample (1).

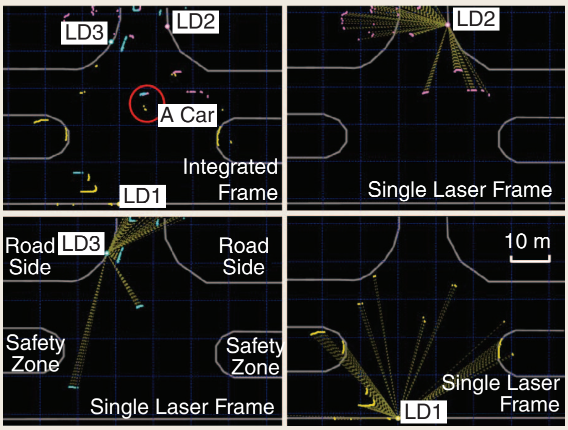

FIG 5 A laser data sample (2).

Let's look at Figure 4 first. An integrated frame that contains the simultaneous laser scans from all three sensors is shown on the top-left. The laser points are displayed in different colors (i.e., pink, yellow and blue), corresponding to different laser scanners. As a comparison, the three individual laser scans are also given and dotted lines are drawn between the sensor and laser points, demonstrating valid return of the laser beams. We found that, by integrating the laser measurements from different viewpoints on the road side, a more complete contour can be recovered for an object that entered the intersection. This means that a more accurate estimation of the object state might be achieved by using such a method of integration. Meanwhile, the objects that were far from one sensor could be caught by another. Such cooperation makes it possible to adequately cover a relatively large intersection.

Let's look at Figure 5 next. More problems can be found here. The laser points in the red circle are an assembly of the measurements to a single object (e.g. a car) from different laser scanners, although they are not spatially connected. From the data appearance and spatial distance, it is difficult to tell that the data are the measurements to the same object. Integrating the data measurements from different sensors certainly provides more possibilities in understanding the objects in an environment, but it also brings many difficulties in data handling: data of the same object might be spatially disconnected. In order to achieve an accurate estimation to the object states, an efficient object model and data association method is required so that to integrate the data measurements of the same object from different sensors into one unit, and to associate the data measurements in-between of different frames to track the objects' motion.

B. Definition to an Object Model

There is a special cue in the laser measurement of an object. This is described in Figure 6. As demonstrated in Figure 6(a), suppose a laser scanner does counterclockwise scanning, a car, which is simplified using a rectangle, is measured by a sequence of laser points from s to e. A corner detector and line fitting is conducted on the laser points so that to detect the edge(s) that represent one or two vertical sides of the car. Along each edge, we define a directional vector u_i to be from a later measured point to an earlier one. No matter where the laser scanner is placed, such a directional vector is unique for each side of the object. This is demonstrated

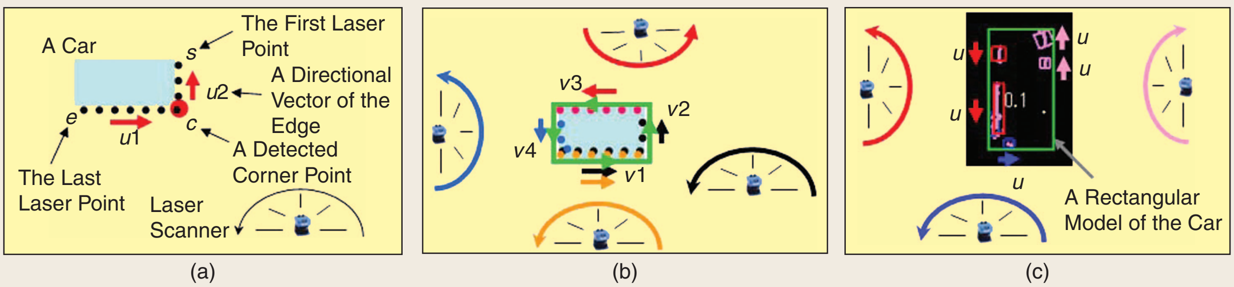

FIG 6 From laser data to an object model.

in Figure 6(b). Suppose v_1, v_2, v_3, v_4 are the directional vectors of the sides of a car, and they compose a counterclockwise loop. A measurement u_i of the car from an arbitrarily located laser scanner should be equal to its corresponding v_j if free of error. Thus, by matching u_i with v_j s, it is easy to find the corresponding side of the measurement, and it is possible to reduce the mismatching based on data appearance. On the other hand, as demonstrated in Figure 6(c), the measurements of a car from different sensors are spatially disconnected, while the u_i s that are extracted from each measurements compose a counterclockwise loop. In this research, an object is simplified using a rectangular model that has a directional vectors along each side, and compose a counterclockwise loop. If the motion direction of the object is detected, v_1 is assigned equal to the motion direction.

Intersections are natural focal points of traffic conflicts. Infrastructure-based monitoring is widely considered as a key technology to relieve these.

C. Moving Object Detection

There are two requirements for the detection of moving objects: to definitively detect all moving objects and to prevent duplicate detections. The first requirement is easily met as well as the moving object reflects the laser beam and it is not too close to the background (static objects). The second requirement is difficult to achieve. Since the object can be measured simultaneously by different laser scanners, many alarms might arise from a single moving object. Moreover, due to the many invalid range returns, the contour points of a single object from one laser scan could be discontinuous, so that a number of detections might arise. Considering these challenges, the detection of moving objects is conducted in this research as follows.

First, on the level of each single laser scan, a segmentation procedure is conducted by examining the continuity in range values, and a rectangular boundary is calculated to fit on each cluster of laser points. Next, a grouping procedure is conducted on the integrated frame level, where, *groups* are extracted by merging nearby *clusters* as follows. At each iteration, a new object is initialized by any of an unassociated cluster. A rectangular boundary is calculated, and a fitness score that represents how well the grouped laser points fit on the estimated rectangular boundary, is estimated for the object. In calculating the fitness score, a projection is made for all the laser points that are grouped inside the object to their nearest edge of the rectangular boundary. The fitness score is defined as the variance of those projection residuals. A small score means a good level of fitness. After initialization, the object tries to absorb other unassociated clusters and/or existing objects. Whenever a candidate is found, which has an overlapped region with the object, a new rectangular boundary, as well as a new fitness score, is calculated by merging them. If the fitness score is above a certain threshold (i.e., worse fitness than a tolerance level), set at 0.3 in this research, the merge is rejected. Conversely, if the fitness score is below 0.3, the merge is accepted.

Figure 7 gives an example of a clustering and grouping result. The rectangular boundary of an object is shown in green, while clusters are shown in different colors corresponding to different laser scanners. It can be noted that each green boundary (object) holds a number of colored boundaries (clusters). The fitness scores are de monstrated in Figure 7 (left). Considering the range error and different object models, normally, a correct estimation will have a fitness score of less than 0.3. For example, the red circle marked with a rectangular boundary in Figure 7 (left) is an erroneous merge of two objects standing closely together. Before the merge, the fitness scores are 0.1 for each (see enlarged subfigure); after the merge, it becomes 0.93. Since this score is higher than our predefined threshold, the merge is rejected.

D. Moving Object Tracking

A crucial procedure in moving object tracking is to associate the data of the same object in different frames. This is more difficult when the measurement data come from a number of distribute sensors. While let's first examine the data from a single laser scanner. Figure 8 (bottom) shows a piece of range image measured by a single laser scanner. Each column represents the range values in one laser scan, each row represents the range returns at the same beam angle along time axis. A static object appears in the range image as a horizontal bar with equal width, while a moving object appears

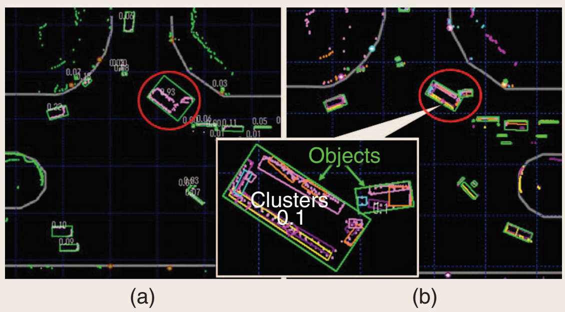

FIG 7 Clustering and grouping the measurements, a correct one (b) comparing to an erroneous one (a).

Experiments in central Beijing show that several laser scanners and a camera can efficiently measure a large deal of detailed trajectory data at intersections.

as a curved band with continuously changing width. As an object passes through the intersection (as well as through the laser scanner), the curved band is always thick in the middle section (many contour points are measured) and thin at the two ends (few contour points are measured). It is obvious that the measurements for a single object, no matter if it is static or mobile, are continuous in both spatial and temporal domains. Figure 8 (up) gives a segmentation result of range points based on the Euclidean distance with their neighbor points. Be notice that totally seven different colors are used to represent all segments in a rotation, some segments are assigned to same colors. It is obvious that by examining spatial and temporal continuity, the "bands" of moving objects can be easily extracted, and, data association in the domain of a single laser scan data can somehow reduce the difficulty.

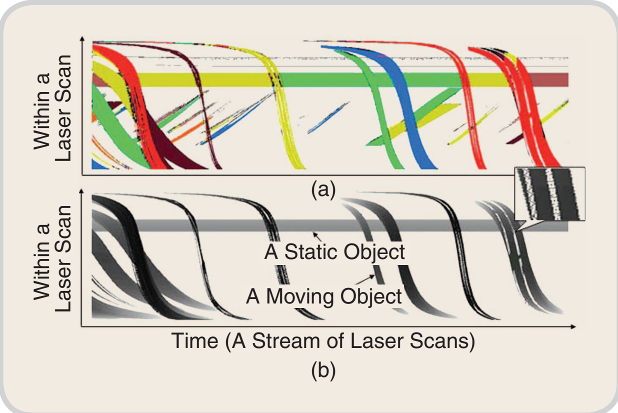

FIG 8 A piece of range image measured by a labeling result (a) and a laser scanner (b).

Our method for tracking moving objects is a bottom-up procedure as shown in Figure 9. A *group* is treated as the observation of a moving object so as to track its state and generate a *trajectory* along time, while data association is conducted mainly in the domain of *clusters*. It is conducted as follows. A *trajectory* is updated in the previous frame $k-1$ by a *group* that is composed of a number of *clusters*. For each *cluster*, in the current frame k of the same laser scanner, we first look at its corresponding(s)according to the data's spatial continuity. Such *clusters* might belong to a number of *groups* in frame k, where the nearest one to the predicted state are used to update the *trajectory*, others are treated as newly appeared objects and new *trajectories* are created accordingly.

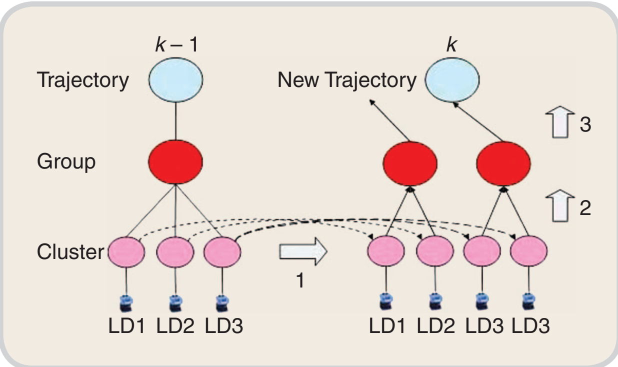

FIG 9 A strategy for moving object's tracking.

IV. Experimental Results

A number of experiments have been conducted in different intersections in central Beijing. Here we present an experimental result that was conducted at a three-way intersection near the campus of Peking University. The intersections in Beijing are famous for their crowdedness. A characteristic of the intersections is that they have many kinds of traffic objects, such as: people, two-wheel bicycles, three-wheel bicycles, motor-bicycles, cars, trucks, and buses. Our goal is to collect the motion trajectories of each individual object and estimate the following parameters: speed, direction, size, class. The many kinds of traffic objects, as well as the many kinds of traffic patterns, in such a crowded environment make this task very challenging.

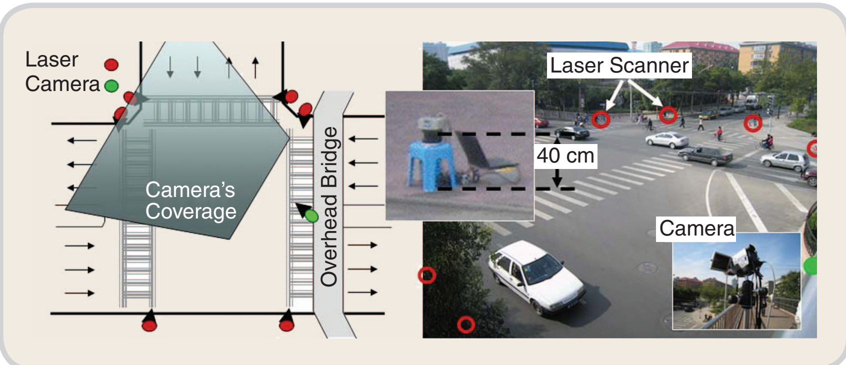

FIG 10 An experiment: sensor layout.

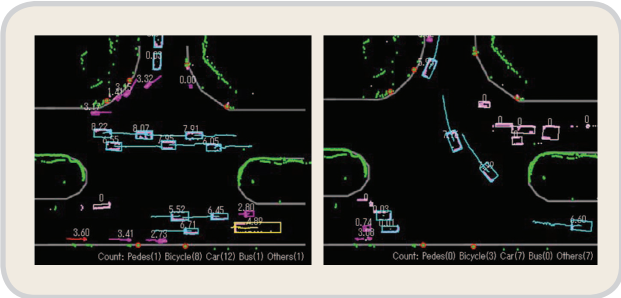

FIG 11 Screen copies of the tracking result.

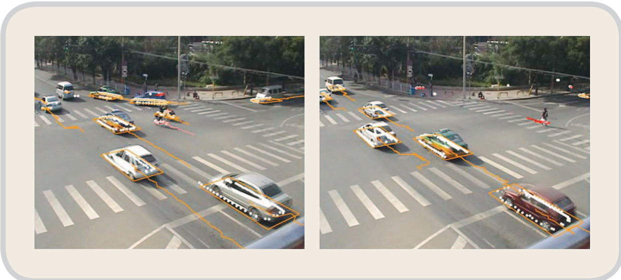

FIG 12 Overlapping of laser-based results onto video image.

Figure 10 shows the sensor layout of the experiment and a picture of the experimental site. Six laser scanners were set on the road side of the intersection. It is better to let the laser scanners profile at the same horizontal plane, but, in a real situation, this requirement is difficult to meet. In this research, each laser scanning plane was set horizontally, at 40cm above the local ground surface. Each laser scanner was controlled by a client computer. The client computers were connected to a server computer through a wireless network. A video camera was also associated, set on an overhead bridge nearby. Currently, the data were mainly used for visualization and result validation, while in future, we are going to develop a fusion-based approach for moving object detection and tracking. In the experiment, due to unstable network condition, data transfer was not conducted on-line. Data from different sensors were integrated and processed in an off-line mode after the experiment.

Figure 11 shows screen copies of moving object detection and tracking using the network of six laser scanners. The figure on each of the object is speed in unit: m/s, while some other states such as size, direction are also estimated. By a threshold-based classification, the objects are divided into people, bicycles, cars and buses according to their size, and demonstrated in red, pink, water blue and yellow respectively. Their counted number at the current frame are also demonstrated at bottom right. On the other hand, the laser-based results are forwarded to a video client, and overlapped onto the video image of the same time stamp. Figure 12 gives two overlapping results, where the estimated rectangular contour of each moving object at the scanning plan is drawn, as well as a 'tail' representing the motion trajectory during the last 20 frames. Different colors represent different objects, i.e. orange for car, pink for bicycle, and red for people.

In this particular experiment, data measurement lasted for half an hour. A great deal of moving objects are detected and their trajectories are output for further analysis. For example, in 20 minutes measurement, totally 2867 trajectories are recorded, among which 1583 are recognized as cars, others are either pedestrians or bicycles. At a certain moment, more than thirty objects can be tracked simultaneously.

V. Conclusion

In this research, we propose a novel system for sensing an intersection using a network of laser scanners and video cameras. This work focus on a laser-based approach on moving object detection and tracking. An experiment in central Beijing is presented, where six laser scanners and a video camera are exploited to cover a three-way intersection. Experimental results demonstrate that such a system can efficiently measure a large deal of detailed trajectory data for further traffic analysis. Future work will be addressed on a fusion-based approach of moving object detection and tracking, and a method of visualizing and analyzing such a large quantitative of traffic data.

References

[1] C. J. Pai, et al., "Pedestrian detection and tracking at crossroads," *Pattern Recognit.*, vol. 37, no. 5, pp. 1025–1034, 2004.
[2] S. Kamijo, Y. Matsushita, K. Ikeuchi, and M. Sakauchi, "Traffic monitoring and accident detection at intersections," *IEEE Trans. Intell. Transport. Syst.*, vol. 1, no. 2, pp. 108–118, 2000.
[3] H. Veeraraghavan, O. Masoud, and N. Papanikolopoulos, "Computer vision algorithms for intersection monitoring," *IEEE Trans. Intell. Transport. Syst.*, vol. 4, no. 2, pp. 78–89, 2003.
[4] D. Streller and K. Dietmayer, "Object tracking and classification using a multiple hypothesis approach," in *Proc. IEEE Intelligent Vehicles Symp.*, 2004, pp. 808–812.
[5] E. Prassler, J. Scholz, and A. Elfes, "Tracking people in a railway station during rush-hour," in *Proc. ICVS99 (LNCS 1542)*, H. I. Christensen, Ed., 1999, pp. 162–179.
[6] H. Zhao and R. Shibasaki, "A novel system for tracking pedestrians using multiple single-row laser range scanners," *IEEE Trans. Syst., Man, Cybern. A*, vol. 35, no. 2, pp. 283–291, 2005.
[7] R. Rad and M. Jamzad, "Real time classification and tracking of multiple vehicles in highways," *Pattern Recognit. Lett.*, vol. 26, pp. 1597–1607, 2005.
[8] S. Atev, H. Arumugam, O. Masoud, and R. Janardan, "A vision-based approach to collision prediction at traffic intersections," *IEEE Trans. Intell. Transport. Syst.*, vol. 6, no. 4, pp. 416–423, 2005.
[9] L. Alexander, et al., "The Minnesota mobile intersection surveillance system," in *Proc. IEEE Intelligent Transportation Systems Conf.*, 2006, pp. 139–144.
[10] A. Mendes and U. Nunes, "Situation-based multi-target detection and tracking with laserscanner in outdoor semi-structured environment," in *Proc. Conf. Intelligent Robots and Systems*, 2004, pp. 88–93.
[11] T. Vu, O. Aycard, and N. Appenrodt, "Online localization and mapping with moving object tracking in dynamic outdoor environment," in *Proc. IEEE Intelligent Vehicle Symp.*, 2007, pp. 190–195.
[12] F. Fayad and V. Cherfaoui, "Tracking objects using a laser scanner in driving situation based on modeling target shape," in *Proc. IEEE Intelligent Vehicle Symp.*, 2007, pp. 44–49.
[13] K. Nakamura, et al., "Tracking pedestrians using multiple single-row laser range scanners and its reliability evaluation," *Syst. Comput. Japan*, vol. 37, no. 7, pp. 1–11, 2006.
[14] H. Zhao, X. W. Shao, K. Katabira, and R. Shibasaki, "Joint tracking and classification of moving objects in cross road using single-row laser range scanner," in *Proc. IEEE Int. Conf. Intelligent Transportation Systems*, 2006, pp. 287–294.

ITS