Proceedings of the
2007 IEEE Intelligent Vehicles Symposium
Istanbul, Turkey, June 13-15, 2007

An efficient extrinsic calibration of a multiple laser scanners and cameras' sensor system on a mobile platform

Huijing Zhao, Yuzhong Chen, Ryosuke Shibasaki

Center for Spatial Information Science
Univ. of Tokyo
chou@iis.u-tokyo.ac.jp

Abstract— **This work is motivated by a development of a portable and low-cost solution for road mapping in downtown area using a number of laser scanners and video cameras that are mounted on an intelligent vehicle. Sensors on the vehicle platform are considered to be removable, so that extrinsic calibrations are required after each sensors' setting up. Extrinsic calibration might always happen at or near measuremental sites, so that the facilities such as specially marked large environment could not be supposed as a given in the process. In this research, we present a practical method for extrinsic calibration of multiple laser scanners and video cameras that are mounted on a vehicle platform. Refering to a fiducial coordinate system on vehicle platform, a constraint between the data of a laser scanner and of a video camera is established. It is solved in an iterative way to find a best solution from the laser scanner and from the video camera to the fiducial coordinate system. On the other hand, all laser scanners and video cameras are calibrated for each laser scanner and video camera pair that has common in feature points in a sequential way. An experiment is conducted using the data measured on a normal street road. Calibration results are demonstrated by fusing the sensor data into a global coordinate system.**

I. INTRODUCTION

In recent years, single-row laser range scanners (briefly called laser scanner) mounted on intelligent vehicles have become very popular for both perception and navigation tasks. They are capable of providing accurate range measurements, in a large angular field, on a certain plane, and with high frame rate. The data size is small and computation cost is always low. It has efficiency in both data acquisition and driving assistance even vehicles run at high speed. On the other hand, video cameras have been studied for decades. They are capable of providing rich sensing data to the environment in the form of intensity images. A great deal of methods have been proposed for acquiring informations from visual data, such as tracing lane marks, detecting road surfaces/road boundaries, tracking moving objects, extracting traffic signs and so on. It has also been demonstrated by a number of existing works that integrating range sensors with video cameras will make the system much powerful [5-9]. For such a purpose, calibration is an essential requirement to represent the different data sources in a common reference frame. Calibration can be decomposed into intrinsic calibration and extrinsic calibration. Intrinsic calibration is to find the parameters such as camera's focal length, optical distortion, principal point and so on. These parameters could be estimated independently with other sensors. Always they have only slight changes with temperature and humidity. There have been dozens of works focusing on the intrinsic calibration of cameras, where Zhang,1999[10] and Tsai,1987[5] are among the most practical methods. Extrinsic calibration is to find the relative rotation and translation between different sensor coordinate systems. These parameters change during sensors' removal and setting up. They have to be calibrated integrated with the whole sensor system. The process of extrinsic calibration of laser scanner and video camera is often poorly documented. In many publications, extrinsic calibration parameters are treated as an existing known by pre-calibration, while few details can be found on how this is done. However, it is an common knowledge that the process of extrinsic calibration is always notoriously labor intensive.

This work is motivated by a development of a vehicle-borne sensing system using a number of laser scanners and video cameras. A picture of the system can be found in Figure 1. The final target of the research is to find a **portable** and **low-cost** solution to road mapping from the streets in downtown areas. For the sake of portability, sensors are assumed to be transported independently with the whole system, so that extrinsic calibrations are required after sensors' setting up. Sensors' setting up is supposed to be conducted near the site of each field survey, so that extrinsic calibration could not assumed any special facility, such as specially marked environments. For the sake of low-cost and privacy issues, it is assumed that laser beams are invisible and the intensity values such as laser beam's reflectance are not able to ease the work in looking for corresponding features of different sensor data. The type of laser scanners used in the system are LMS200 and LMS291 by SICK AG.

To be brief, this research is objective to develop a fast and cheap extrinsic calibration method, so that the laser points from different single-row laser range scanners could be integrated into a common reference frame, and they are able to be back-projected onto the video images that are captured by different video cameras for further analysis. Requirements to the calibration method are, 1) it do not rely on any special facilities; 2) it is a fast and cheap solution; 3) it is desirable that the calibration could be conducted on street and with few labor work at the field.

We have found three related works addressing the extrinsic calibration of laser and cameras. Unnikrishnan and Hebert, 2005[1] developed a very practical method of **fast** extrinsic calibrating a **3D** laser scanner and a video camera by using a portable checkerboard. The checkerboard is placed at a number of poses where both laser and camera could clearly see it. The planar surface of the checkerboard at each pose are calculated from each data. They are matched to find a best solution to the rotation and translation between two sensor's coordinate system. The method is difficult to be applied to our case as it is impossible to calculate the planar surface from an instantaneous measurement of a single-row (i.e.**2D**) laser range scanner. Mertz, et al. 2002[8] calibrated a laser line stripper with a video camera, however one of the basic requirement is that laser line is visible on video image. Zhang and Pless, 2004[2] proposed an extrinsic calibration of a **2D** laser scanner and a video camera that are mounted on a robot platform. A checkerboard is also used in the research. The planar surface of the checkerboard is calculated from the video image. Rotation and translation between laser scanner and video camera is solved to meet the contraint that laser points fit on the planar surface. In fact we do not believe that the method will work. A 2D laser scan cut through a line on the checkerboard. However there are no guarantee in the constraint which line of the planar surface should the laser points fit on.

One of the major keys to a reliable calibration is to find enough number of corresponding features. A common way is to locate a certain calibration pattern at different poses, Like [1] and [2] did. This has efficiency when the measurement field is at a low elevation, so that the calibration pattern could be made in a portable size, and located easily, e.g. put on the ground. However if the laser beam scans upwardly or vertically, it will be difficult to catch the laser beams at high elevations by using the calibration patterns. In addition, as has been addressed before, a 2D laser scan cut through a line on a surface in real world, while it is an uncertainty which line it went through. It is difficult to establish an exact correcpondence between the data of laser scan and video image, if neither assume sensors' setting condition, e.g. horizontal scanning at a certain elevation, nor do any physical measurement to locate the laser beams. When the laser beam scans horizontally, this is easy to be solved by assuming the height of laser points are know, as is conducted in our previous research[3]. In this research, we focus on the calibration when laser scan plane is not horizontal.

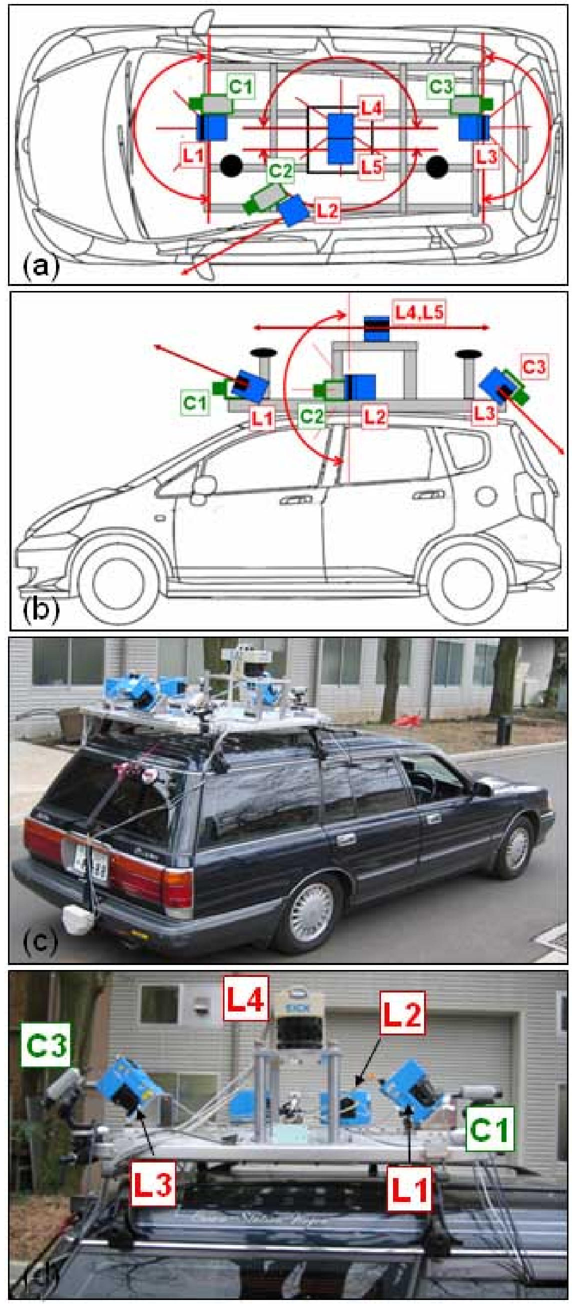

Fig. 1. Configuration of laser scanners and video cameras on the vehicle-borne sensor syste

The way we took in this research is to let the sensors' platform move. As sensors are fixed on the platform rigidly, relative transformations, i.e. extrinsic calibration parameters, between each sensors keep the same. Along with the platform's motion, 2D laser scanning could compose a 3D observation to the surroundings. A geometric feature in laser view could be captured in a number of video images from different viewpoints. The few number of geometric features in laser view could be associated to many times of image points along a video stream so that the number of corresponding features could be easily increased.

The paper is organized as follows. Section2 briefly describe the vehicle-borne sensor system. Section3 outlines the methods for calibration. Section 4 gives experimental results, where totally three laser scanners and three video cameras on the vehicle-borne system are calibrated using a data that are measured on a normal street road.

II. VEHICLE-BORNE SENSOR SYSTEM

Figure 1 outlines the configuration of laser scanners and video cameras in the vehicle-borne sensor system. Sensors are laid out on two stories. Although a GPS/IMU/VMS based navigation unit is implemented to acquire position and orientation data of the vehicle platform at each time interval, its accuracy is not enough for detailed 3D mapping in many downtown areas due to the always block of GPS signal and error accumulation in IMU. Two laser scanners, L5 and L6, covering horizontal angular field of 360 degree at the top story, are exploited to assist for a better navigation accuracy. (A detailed description to the laser scanner based navigation for detailed 3D mapping can be found in [4]). Three pairs of laser scanner and video camera are mounted on the lower story of the platform, facing to different targets. The first pair, L1 and C1, face to the front. L1 scans upwardly from the right to the left. Reason for L1's setting is to get a dense point sampling to the vertical features, such as traffic signals, signboards, building edges and so on. Objects on both sides of the road could be captured simultaneously by L1, if their elevation are higher than L1's mounting position. The second pair, L2 and C2, face to the front-left. (In Japan, cars run on the left side of roads). L2 does vertical scanning in order to see the objects on left, such as road boundaries, poles, side traffic signboards, buildings and so on. The third pair, L3 and C3, face to ground surface and mounted on the rear of the vehicle. L3 does downward scanning in order to acquire a dense point sampling to road surface and road boundary. Objects below L3's mounting position could be measured. There were originally a fourth pair mounted on the vehicle, facing to the right-left. Their function are almost the same with L2 and C2, while their video images are not as good as the others that have longer vision depth. They are not being used now, so are not subject to calibration. Video cameras used in the system are Sony's DFW-SX910 with wide-angle lens.

III. EXTRINSIC CALIBRATION OF MULTIPLE LASER SCANNERS AND VIDEO CAMERAS

In this research, we pre-calibrate the intrinsic parameters of video camera by using Tsai's model[5], and assume that laser scanner has a white noise in range distances, while no systematic error in scanning angle. This document concentrate on the extrinsic calibration only, i.e. to find the relative rotation and translation between different sensor coordinate systems. There are always two basic requirements when fusing the different sensor data from laser scanners and video cameras: 1) projecting laser points to a video image ; 2) integrating the laser points from different laser scanners to a common reference frame. In this research, we define a fiducial coordinate system, called vehicle's body coordinate system, and relate each sensor's coordinate system by establishing constraints from each laser scanner and each video camera to the fiducial coordinate system. Figure 3 and 4 give conceptual images of fusing the data from different laser scanners and video cameras, where blue words denote that the transformation parameters here are considered as given, while red words denote for those to be calculated in this research.

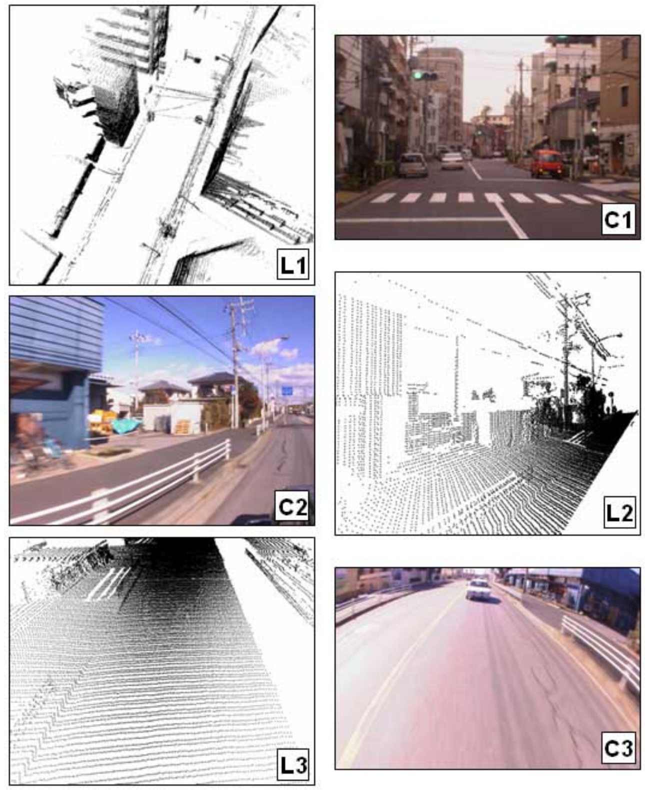

Fig. 2. Example of the data from each laser scanner and video camera

Without loss of generality, we define laser scanner's coordinate system as follows. Origin is at the principle point of laser scanner. Laser beam scans starting from X-axis towards Y-axis. Z-axis is vertical to the laser beams scanning plane, and XYZ axes compose a right-hand coordinate system. A laser point P_l is measured by laser scanner l at time k_l with a range distance of r at the angle of α. P_l at laser scanner l's coordinate system is formulated as follows.

$$P_l = (-r \sin \alpha, 0, -r \cos \alpha, 1)^t \quad (1)$$

A vehicle's body coordinate system and a global coordinate system are defined according to the navigation mechanism, where the navigation output associate vehicle body to global coordinate system. In calibration stage, it is not necessary to geo-reference the vehicle's position and orientation to any world coordinate syste. The global coordinate system could be defined to any relative reference frame. A basic requirement is that motion of vehicle body in the reference frame could be accurately located. Let T_{vw}^{kl} be the transformation matrix from vehicle's body coordinate system to the global coordinate system that output from navigation unit at time k_l, T_{lv} be the transformation matrix from laser scanner l to vehicle's body. Laser point P_l could be associated to the global coordinate system as follows.

$$P_w = (x_w, y_w, z_w, 1)^t = T_{vw}^{kl} \cdot T_{lv} \cdot P_l \quad (2)$$

In this document, notation T_{ij}^{k} means a transformation matrix from the coordinate system i to j at time k. Transformation matrixes T is formulated as follows.

$$T_{4*4} = \begin{pmatrix} R_{3*3} & -\Delta_{3*1} \\ 0 & 1 \end{pmatrix}$$

On the other hand, a video image I is captured by video camera c at time k_c. Let $P_i = (x_i, y_i, z_i, 1)^t$ be the coordinates of P_l at camera c's coordinate system, p_i be an image point of P_l on video image I. A constraint between laser and image points is built as follows.

$$P_i = T_{vc} \cdot T_{vw}^{kc} \cdot T_{vw}^{kl} \cdot T_{lv} \cdot P_l \quad (3)$$

$$p_i \sim K(P_i) \quad (4)$$

where, $T_{wv}^{kc} = {T_{vw}^{kc}}^{-1}$ is the transformation matrix from the global coordinate system to vehicle's body coordinate system that output from the navigation unit at time k_c. $T_{vc} = T_{cv}^{-1}$ is the transformation matrix from vehicle's body coordinate system to camera c's coordinate system, K is a perspective transformation from camera's coordinate system to image plane, which is defined by camera's intrinsic calibration parameters.

T_{wv}^{kc} and T_{vw}^{kl} in Equation (3) are supposed to be accurate enough in extrinsic calibration. If an expensive IMU and/or GPS is not available in use, horizontally scanning laser scanner could be a very cheap solution, but provide good results for calibration, as what we did in experiments. An algorithm of positioning using horizontally scanning laser scanner and some experimental examination can be found in [4]. A limitation of the method is that it could not detect vehicle's vertical motion, so that the running course must be flat. Given a number of corresponding $(< P_l, k_l >, < p_i, k_c >)$ pairs, the objective is to find a solution to the unknowns T_{vc} and T_{lv} in Equation (3), which are the extrinsic calibration parameters relating camera and laser scanner to the vehicle's body coordinate system. This could be solved in an iterative way. In each iteration, either T_{vc} or T_{lv} are treated as a known, while looking for an answer to the other one. The process continues until it convergence.

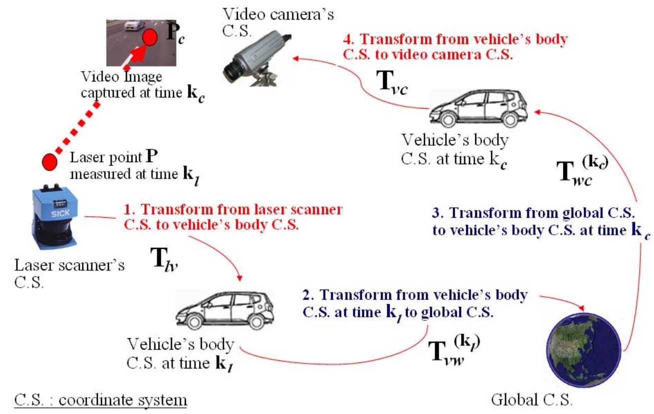

Fig. 3. Procedure of projecting a laser point onto a video image

Suppose T_{lv} is known, for each $(< P_l, k_l >, < p_i, k_c >)$ pair, the navigation out at time k_l and k_c are found out, P_v, the coordinates of P_l at the vehicle's body coordinate system at time k_c, is calculated as follows.

$$P_v = T_{wv}^{kc} \cdot T_{vw}^{kl} \cdot T_{lv} \cdot P_l \quad (5)$$

So that each $(< P_l, k_l >, < p_i, k_c >)$ pair is converted to $(< P_v, k_c >, < p_i, k_c >)$, where the relation left between P_v and p_i is the transformation from vehicle's body coordinate system to camera c's image plane.

$$p_i \sim K(T_{vc} \cdot P_v) \quad (6)$$

It is well studied problem in the field of computer vision, and could be solved by many existing methods, e.g. Tsai [].

Now let's consider the case that T_{vc} is known. Let $p_u = (u, v)^t$ be the undistorted coordinates of p_i after an inversed conversion on camera's distortion parameters, (u_0, v_0) principle point of image plane, d focal length. Calculoate $P_u = (u - u_0, v - v_0, -d)^t$ for each p_i, where $P_u = c \cdot P_i$, c is an unknown scaling factor. For each $(< P_l, k_l >, < p_i, k_c >)$ pair, a $P_u = (u - u_0, v - v_0, -d)^t$ is first calculated on p_i, then converted to a pseudo perspective space at the vehicle's body coordinate system at time k_l in order to remove the time factor among them.

$$P'_v = T_{wv}^{kl} \cdot T_{vw}^{kc} \cdot T_{cv} \cdot P_u \quad (7)$$

where,

$$P'_v = c \cdot T_{lv} \cdot P_l \quad (8)$$

relating the coordinate systems between vehicle's body and laser scanner l only. T_{lv} can be solved as a direct linear transformation (DLT) problem.

IV. EXPERIMENTAL RESULTS

An experiment is conducted at a normal street road in downtown area. A 2D map and vehicle's motion trajectory is shown in Figure 5, which is generated by matching the horizontally scanning laser range data from

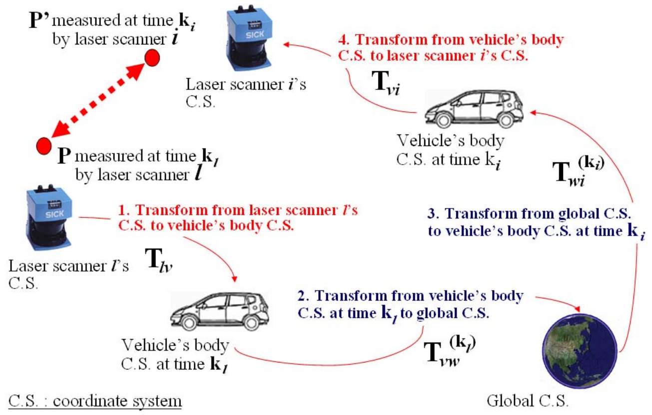

Fig. 4. Procedure of projecting a laser point measured by one laser scanner to the coordinate system of another laser scanner

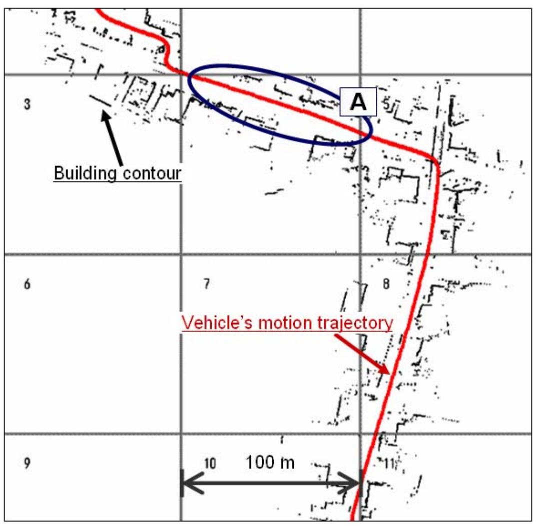

Fig. 5. A 2D map and vehicle's motion trajectory is generated by the horizontal laser scanning of L4 and L5. Data in region "A" is used for calibration.

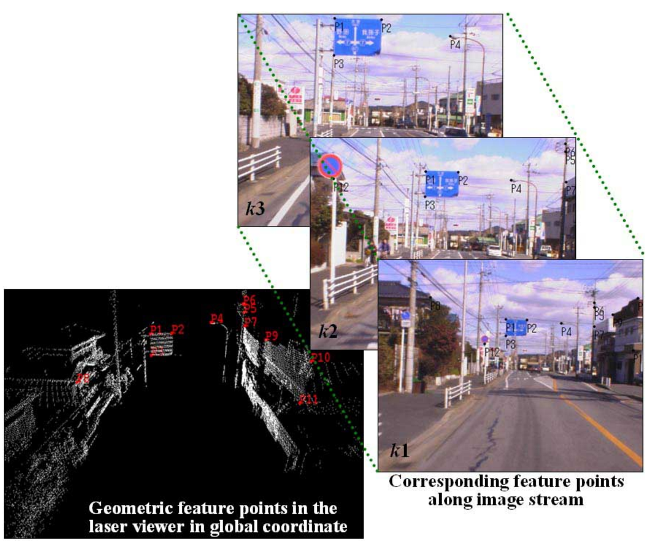

Fig. 6. Procedure of projecting a laser point measured by one laser scanner to the coordinate system of another laser scanner

L4 and L5. Both map and motion trajectory are relative to a reference coordinate system, which coincide with the vehicle's body coordinate system at the starting point. The data in region "A" are used for calibration, more specifically, taking corresponding feature points, as the road surface in region "A" is almost flat (see Fig 2), and the vehicle ran through it smoothly without any sudden speed change. Corresponding feature points from different data sets are picked up manually in this experiment. Figure 6 shows an example of this process when calibration laser scanner L1 and video camera C1. The left is a 3D view of the laser points P_ws from L1, which are integrated to the global coordinate system on Equation (2). Here T_{lv} is assigned according to the hardware specification. It is applied as the initial value in calibration process. Feature points P_ws are picked up manually in this view, while their corresponding $< P_l, k_l >$ are recorded. The right is a sequence of video images from C1. For each P_w, a p_i is first manually picked up on a certain video image k_c, then other $< p_i', k_c' >$ pairs are looked along the image sequence by image matching with the template near $< p_i, k_c >$. It is not necessary to find all them, while it is suggested that p_is cover a larger area on image plane but k_c distribute in a shorter time strip to minimize the influence from navigation error. A total of 30 $\sim$ 40 corresponding feature pairs, $< P_l, k_l, p_i, k_c >$ are collected for calibration of each pair of laser scanner and video cameras. Figure 7 show a result of projecting the laser points by L1, L2, L3 to a video image by C2. In addition, by calibrating two laser scanners with a common video camera, mismatching between the laser points from different two laser scanners could be reduced. This is demonstrated in Figure 8. So in this experiment, laser scanners and video cameras are calibrated in the following sequences.

1) Initialize T_{lv2} according to the hardware specification
2) Calibrate L2 and C2 to refine T_{lv2} and obtain T_{cv2}
3) Let T_{cv2} be a known, calibrate L1 and C2 to obtain T_{lv1}
4) Let T_{lv1} be a known, calibrate L1 and C1 to obtain T_{cv1}
5) Let T_{cv2} be a known, calibrate L3 and C2 to obtain T_{lv3}
6) Let T_{lv3} be a known, calibrate L1 and C1 to obtain T_{cv3}.

V. CONCLUSION AND FUTURE STUDY

This work is motivated by a development of a portable and low-cost solution for road mapping in downtown area using a number of laser scanners and video cameras that are mounted on an intelligent vehicle. Sensors on the vehicle platform are considered to be removable, so that extrinsic calibrations are required after each sensors' setting up. Extrinsic calibration might always happen at or near measuremental sites, so that the facilities such as specially marked large environment could not

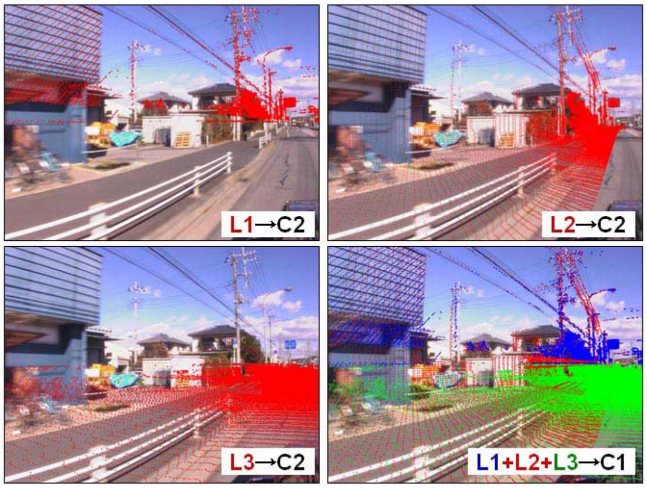

Fig. 7. projecting the laser points by L1, L2, L3 to a video image by C2

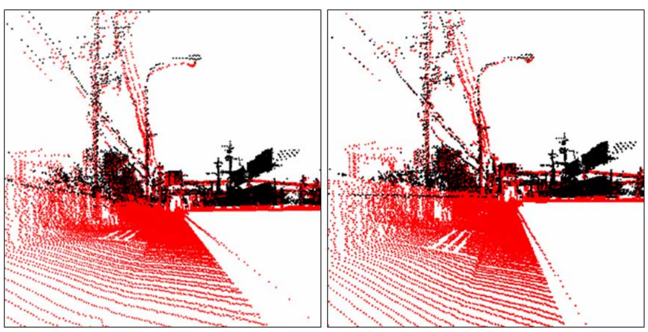

Fig. 8. Mismatching between the data of different laser scanner is reduced after calibration

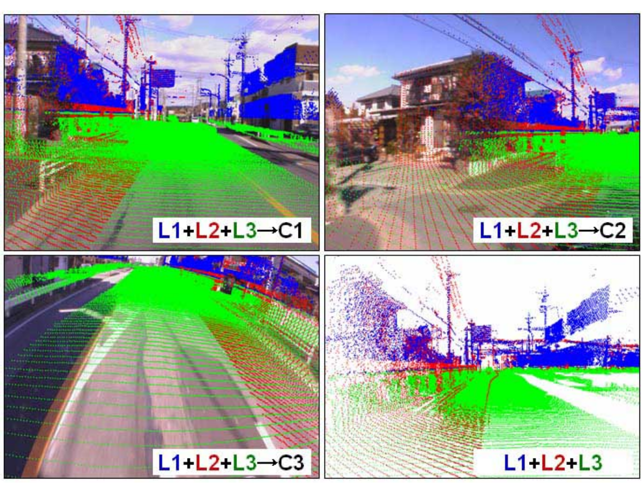

Fig. 9. Final results of calibrating multiple laser scanners and video cameras

be supposed as a given in the process. In this research, we present a practical method for extrinsic calibration of multiple laser scanners and video cameras that are mounted on a vehicle platform. Refering to a fiducial coordinate system on vehicle platform, a constraint between the data of a laser scanner and of a video camera is established. It is solved in an iterative way to find a best solution from the laser scanner and from the video camera to the fiducial coordinate system. On the other hand, all laser scanners and video cameras are calibrated by calibrating each laser scanner and video camera pair that has common in feature points in a sequential way. An experiment is conducted using the data measured on a normal street road. Calibration results are demonstrated by fusing the sensor data.

Future study will focus on reducing the manual work in taking corresponding feature points, and developing a method of simultaneously matching the corresponding features from all sensoring data.

References

[1] Unnikrishnan, R., Hebert, M., Fast Extrinsic Calibration of a Laser Rangefinder to a Camera, *CMU-RI-TR-05-09*, July, 2005.

[2] Zhang,Q., Pless, R., Extrinsic Calibration of a Camera and Laser Range Finder, *Proc. of IEEE Intl. Conference on Intelligent Robots and Systems (IROS)*, vol.3, pp.2301-2306, 2004.

[3] Zhao, H., Shibasaki, R., A real-time system for monitoring pedestrians, *Proc. of 7th IEEE Workshop on Applications of Computer Vision*, 2005.

[4] Tsai, R.Y., A versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses, *IEEE Journal of Robotics and Automation*, Vol. RA-3, No. 4, August 1987, pages 323-344.

[5] Perrollaz, M., et al., Long Range Obstacle Detection using Laser Scanner and Stereovision, *Proc. of IEEE Intelligent Vehicles Symposium (IV'06)* , pp.182-187, 2006.

[6] Xiang Z., Ozguner, U., Environmental Perception and Multi-sensor Data Fusion for Off-road Autonomous Vehicles, *Proc. of IEEE Conf. on Intelligent Transportation Systems (ITSC'05)* , pp.584-589, 2005.

[7] Fruh, C. and Zakhor, A., An Automated Method for Large-Scale, Ground-Based City Model Acquisition, *International Journal of Computer Vision* , Vol. 60, No. 1, pp. 5 - 24, 2004.

[8] Mertz, C., Kozar, J., Miller, J.R., Thorpe, C., Eye-safe Laser Line Striper for Outside Use, *Proc. of IEEE Intelligent Vehicle Symposium*, 2002.

[9] Aufrere, R., Mertz, C., Thorpe, C., Multiple Sensor Fusion for Detecting Locatio of Curbs, Walls, and Barrers, *Proc. of IEEE Intelligent Vehicle Symposium*, 2003.

[10] Zhang, Z., Flexible Camera Calibration by Viewing a Plane from Unknown Orientations, *Proc. of* , 1999.