2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

Accurate Motion Estimation and High-Precision 3D Reconstruction by Sensor Fusion

Yunsu Bok, Youngbae Hwang, and In So Kweon

Abstract—**The CCD camera and the 2D laser range finder are widely used for motion estimation and 3D reconstruction. With their own strengths and weaknesses, low-level fusion of these two sensors complements each other. We combine these two sensors to perform motion estimation and 3D reconstruction simultaneously and precisely. We develop a motion estimation scheme appropriate for this sensor system. In the proposed method, the motion between two frames is estimated using three points among the scan data, and refined by nonlinear optimization. We validate the accuracy of the proposed method using real images. The results show that the proposed system is a practical solution for motion estimation as well as for 3D reconstruction.**

I. Introduction

MOTION estimation and 3D reconstruction are fundamental problems in computer vision and robotics. The most popular sensors are CCD cameras and laser range finders. The CCD camera provides the projected image of an entire 3D scene. However, we cannot obtain the depth information of the scene from an image without constraints. The 2D laser range finder provides the depth information of the scanning plane. However, we cannot obtain the depth information of the entire 3D structure from scan data without an additional device such as a tilting module.

Several methods have been proposed to estimate motion in 3-D space using CCD cameras. The 5-point algorithm [1] and 3-point algorithm [2][9][10] estimate the initial motion using some point correspondences and minimize the reprojection error. Probabilistic approaches based on the extended Kalman filter (EKF) or particle filter provide good motion estimation results [3]. Calibration methods using constraints also have been proposed [16]. However, the 3D reconstruction results are not very accurate because of the limitation of image resolution.

To obtain and utilize accurate depth information, laser sensors can be used. A method for SLAM (simultaneous localization and mapping) using a 2D laser sensor is proposed in [4]. This method requires the 2D laser sensor to be tilted slowly.

A method for using both sensors, camera and laser, is proposed in [5]. A 3D laser sensor is hung under a balloon to scan the upper part of ruins. A camera is attached to the laser sensor to refine the distorted scan data due to the motion of the balloon. The motion of the balloon is estimated using the image sequences captured by the camera, and refined using several constraints from the camera motion and the laser data. Both sensors are also used in [11]. In an indoor environment, a robot's motion is estimated using a 2D laser sensor. The image is then transformed based on this result to facilitate feature matching.

As mentioned above, cameras and laser sensors have different characteristics. If two sensors are combined, their weaknesses can be complemented by each other; e.g. the unknown scale of the single-camera-based approaches can be estimated by scan data, and the 3D motion of a 2D laser sensor also can be estimated by images.

We present a new sensor system, the combination of a camera and a 2D laser sensor. We also present a noble motion estimation method appropriate for this system. The motion estimation and 3D reconstruction are achieved simultaneously using the proposed sensor system.

This paper is organized as follows: Section 2 introduces the proposed sensor system. Section 3 presents the proposed motion estimation algorithm for the system. Experimental results are given in Section 4.

II. Combination of Camera and Laser

A. A New Sensor System

We propose a new sensor system. The sensor system consists of a CCD camera and a 2D laser range finder. We attach a camera at the top of a laser sensor, as shown in Fig. 1. The following algorithms are independent of sensor configuration, but the sensors need to be pointed in the same direction because the scanned points need to be seen by the camera. The relative pose between the sensors is assumed to be fixed.

B. Extrinsic Calibration

To use images and range data simultaneously, it is necessary to transform data into a common coordinate system. Therefore, the relative pose between the sensors should be computed. A method of extrinsic calibration between a camera and a 2D laser sensor was proposed in [6]. The camera is calibrated first using a pattern plane [7] with concentric circles [8]. Then the position of the plane in the camera coordinate system is computed. While the camera

Yunsu Bok is a Ph. D. student of Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea (corresponding author to provide phone: +82-42-869-5465; e-mail: ysbok@rcv.kaist.ac.kr).

Youngbae Hwang is a Ph. D. student of Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea (e-mail: unicorn@rcv.kaist.ac.kr).

In So Kweon is a professor of Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea. (e-mail: iskweon@kaist.ac.kr).

1-4244-0602-1/07/$20.00 ©2007 IEEE.

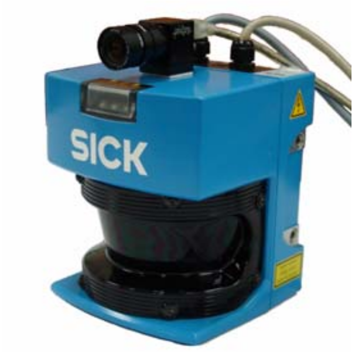


Fig. 1. Sensor configuration

captures the images of the plane, the laser sensor also scans the same plane. The relative pose between the sensors can be estimated because some points of the scanned data are on the pattern plane.

III. Motion Estimation

A. Basic Idea

The basic concept of motion estimation for the proposed sensor system is 3D-to-2D matching. This means that we minimize the projection error of 3D points on the image plane. The most important feature of the system is that the laser data are used as the 3D points. The 3D points obtained from a single or stereo camera system are inaccurate due to the limitation of image resolution. On the contrary, laser points have a very small noise, which is independent of the measured range. (For example, the sensor used in this paper has ±10 mm error at all distances.) This configuration is equal to that of the conventional 3-point algorithm, but the measured 3D points are very accurate.

B. Correspondence Search

To find the location of the current scan data in the following image, we project the current scan data onto the current image and find the corresponding point in the following image, as shown in Fig. 2. Several methods for image matching have been proposed in the literature. Template matching, KLT [13] and descriptor matching of features such as SIFT [12] and G-RIF [15], are good examples. In this paper, we use the KLT tracker.

C. Initial Solution

A degeneracy exists in the conventional 3-point algorithm because of the use of a 2D laser sensor. If the laser sensor scans a plane, the scanned data lie on a line. To avoid this degeneracy, we use the laser data from both frames.

We have two frames of data, and each frame consists of an image and range data. Three laser points are selected from two frames. In Fig. 3, Q_1 and Q_2 are from frame 1, and Q_3 is from frame 2. Transforming them into their own camera coordinates gives us 3D coordinates of each point, and then

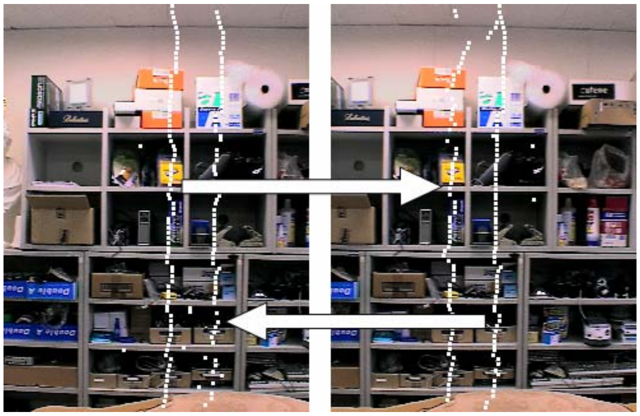


Fig. 2. Example of correspondence search

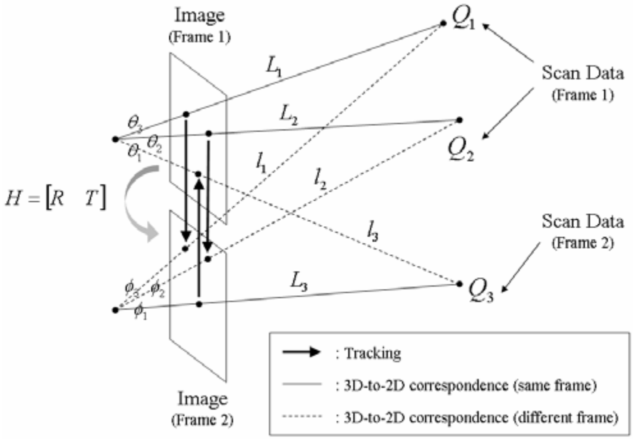


Fig. 3. Initial solution for the proposed sensor system

L_1, L_2 and L_3 are computed. The angles between the camera rays, θ_1, θ_2, θ_3, ϕ_1, ϕ_2 and ϕ_3, are also computed if we find the corresponding image points of projected laser points. (Solid lines and dotted lines represent the 3D-to-2D correspondences; the former shows the correspondence between images and scan data of the same frame while the latter shows that of the other frame.) The unknown lengths are l_1, l_2 and l_3. Applying the second law of cosines, (1), (2) and (3) are derived from Fig. 3.

$$L_1^{\ 2} + L_2^{\ 2} - 2L_1L_2 \cos\theta_3 = l_1^{\ 2} + l_2^{\ 2} - 2l_1l_2 \cos\phi_3 \quad (1)$$

$$L_1^{\ 2} + l_3^{\ 2} - 2L_1l_3 \cos\theta_2 = l_1^{\ 2} + L_3^{\ 2} - 2l_1L_3 \cos\phi_2 \quad (2)$$

$$L_2^{\ 2} + l_3^{\ 2} - 2L_2l_3 \cos\theta_1 = l_2^{\ 2} + L_3^{\ 2} - 2l_2L_3 \cos\phi_1 \quad (3)$$

Solving the equations, we know the coordinates of Q_1, Q_2 and Q_3 in each camera coordinate system. The motion between the frames is computed using these points.

D. Nonlinear Optimization

We refine the initial solution by the nonlinear optimization of a cost function:

$$\sum \frac{\left(q'^{T} Eq\right)^2}{\left(e_1 q\right)^2 + \left(e_2 q\right)^2} + \sum \left(p - \frac{[R \mid T]P}{[r_3 \mid t_3]P} \right)^2 \tag{4}$$

$$[R \mid T] = \begin{bmatrix} r_1 & t_1 \\ r_2 & t_2 \\ r_3 & t_3 \end{bmatrix} \quad E = [T]_\times R = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} \tag{5}$$

where q and q' are the corresponding feature points on the images. P and p are the laser point and its correspondence in the other image. The first term of (4) is the distance between the epipolar line and the corresponding point on the image. The second term is the projection error onto the image. R and T are the rotation and translation matrices, and E is the essential matrix. To reduce the ambiguity due to the narrow field of view, we use corner features extracted by the Harris operation [14].

E. 3D Reconstruction

The 3D reconstruction problem is easily solved with the proposed sensor system. If the motion of the system is estimated, the scan data are transformed into common coordinates. In (6), the rotation R_i and translation T_i are the motion from the *i*-th frame to the (*i*+1)-th frame. The scan data L_i of *i*-th frame are transformed to the data l_{1i} in the laser coordinate system of the first frame. In addition, the texture mapping is easy because all of the scan data are projected onto the images.

$$l_{1n} = \begin{bmatrix} R_1 & T_1 \\ 0 & 1 \end{bmatrix}^{-1} \cdots \begin{bmatrix} R_{n-1} & T_{n-1} \\ 0 & 1 \end{bmatrix}^{-1} L_n = \left(\prod_{i=1}^{n-1} \begin{bmatrix} R_i & T_i \\ 0 & 1 \end{bmatrix}^{-1} \right) L_n \tag{6}$$

Feature points of the images can also be added to the 3D points. If the matching is correct, triangulation generates the 3D information of a wider range than the laser data. However, the 3D points generated by the feature points have large uncertainty in their positions along the viewing direction of the camera. If reducing this uncertainty is impossible using many images, only the laser points should be included in the 3D reconstruction.

F. Algorithm Accuracy

To validate the proposed motion estimation algorithm in a degenerate case, we compare the accuracy of the algorithm to conventional algorithms, perspective 3-point [9] and generalized 3-point [10], using synthetic data. We generate 1,000 motions and 240 3D points similar to the laser data, assuming that the laser sensor scans a plane. Three points are selected randomly 100 times and supplied to the algorithms estimating motion. The Euclidean distance of the translation error is computed as the motion error. To make the synthetic data more realistic, we add Gaussian noise (mean = 0, variance = 15) to the laser data. (This noise is similar to the noise in real sensor data.) For the first experiment, we add round-off noise and Gaussian noise to the images. The result is shown in Fig. 4. In this case, the proposed algorithm works much better than the other algorithms because it is designed to avoid this kind of degenerate case. In the second experiment, we also generate 3D points like the second experiment, but we adjust the distance between the laser sensor and the plane. Fig. 5 is the result with round-off noise on the image. These experiments using synthetic data show that the proposed algorithm proved to be more reliable than the conventional algorithms as the initial motion estimator.

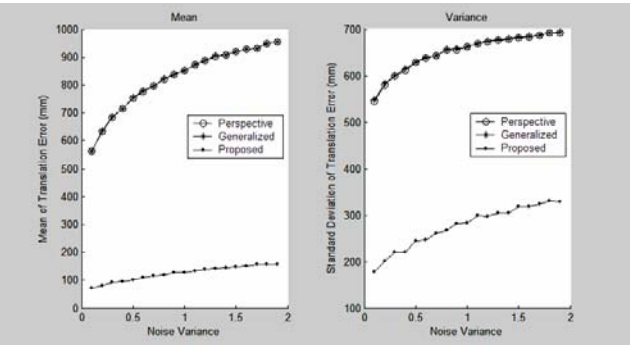


Fig. 4. Motion error using laser points (Noise variance vs. Error) Round-off noise and Gaussian noise are added to the image.

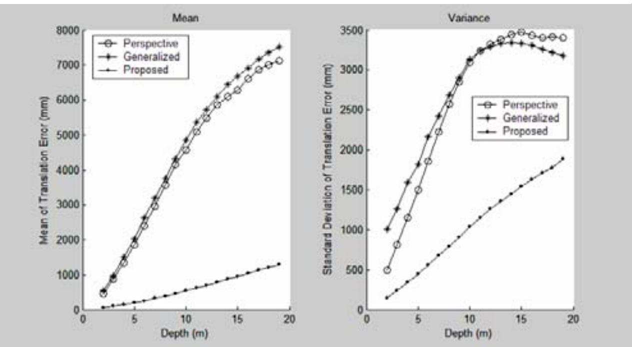


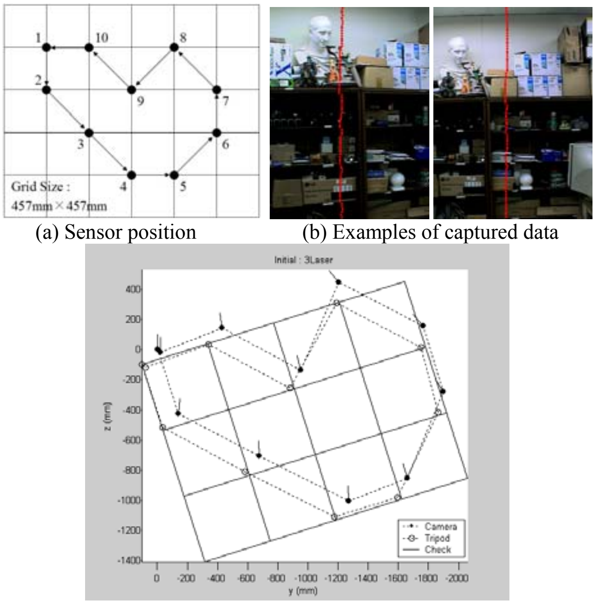
Fig. 5. Motion error using laser points (Plane depth vs. Error) Round-off noise is added to the image.

IV. Experimental Results

A. Motion Estimation

We fix the sensors on a tripod and move the tripod as in Fig. 6(a). The check pattern in Fig. 6(a) is placed on the floor. Several examples of captured data are shown in Fig. 6(b). Fig. 6(c) shows the result of motion estimation. The total length is about 5.7 m and the resulting error using the proposed method is less than 50 mm, as shown in Table I.

To compare results, we perform the same experiment using a stereo camera whose baseline is about 330 mm. We extract the 3-D coordinates of the image point correspondences by triangulation and estimate the initial motion using the ICP [17] algorithm. The initial solution is refined to minimize the projection error. The distances between the true position and estimated position of the tripod are displayed in Fig. 7. The results show that the algorithm using a camera and laser sensor is better than the algorithm using stereo cameras.



(a) Sensor position

(b) Examples of captured data

(c) Motion estimation result

Fig. 6. Motion estimation in indoor environment
(Red: projected laser data, Image size: 320×240)

TABLE I
INDOOR MOTION ESTIMATION ERROR

Frame	Length	Error	Error rate
10	5.7m	50mm	0.9%

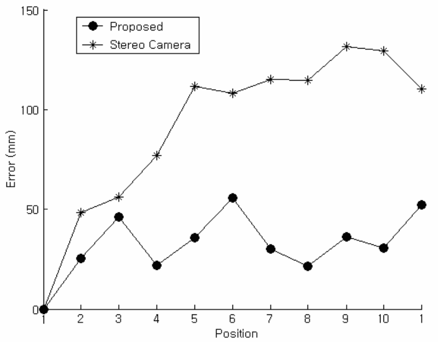
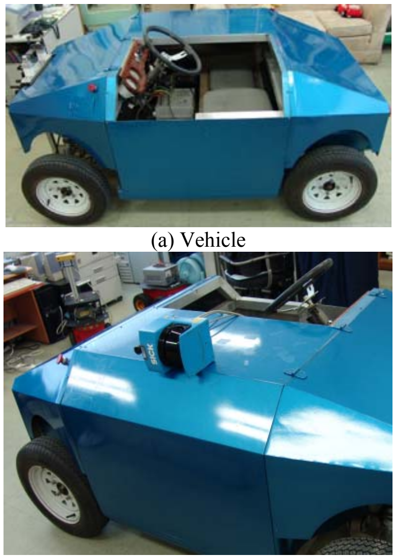


Fig. 7. Position error compared to a stereo based method

B. *Outdoor 3D reconstruction*

We attach the proposed sensor system to the side of an outdoor vehicle shown in Fig. 8(a). The sensor is rotated 90 degrees to scan vertically and pointed to the side of the vehicle, as shown in Fig. 8(b). The data are obtained in the environment shown in Fig. 9, and several examples of captured data are given in Fig. 10. The results of motion estimation and 3D reconstruction are shown in Fig. 11. The black points are the location of the sensor system in each frame, and gray points are the scanned points transformed into the first frame's camera coordinate system. The laser data include the wall of the building and a part of the ground. To verify the accuracy of the results, we extract the wall part from the reconstruction result and compare it to the floor plan of the building. The result in Fig. 12 shows that the 3D reconstruction result overlaps the floor plan very well. The length of the path is about 110 m and the closed-loop translation error is 3.68 m, as shown in Table II. For realistic reconstruction results, we can add texture onto the structure. Fig. 13 shows several parts of the result in Fig. 11 with texture mapping.



(a) Vehicle

(b) The proposed sensor system attached to the vehicle

Fig. 8. Vehicle for outdoor experiment

Fig. 9. Outdoor environment
(The white points on the ground have no relation with this paper.)

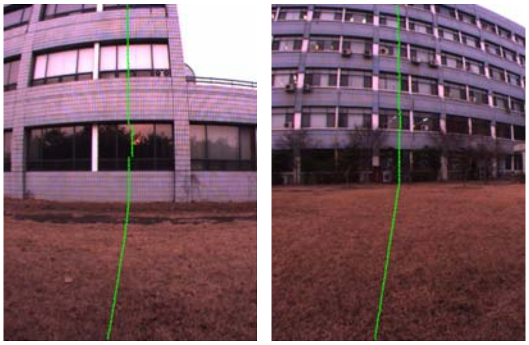


Fig. 10. Example of data captured in outdoor environment
(Green: projected laser data, Image size: 640×480)

C. *Error Analysis*

To compute the accumulated error using real data with unknown true motion, we estimate motion using a part of the sequence used in Subsection 4.B in both directions. If a part is

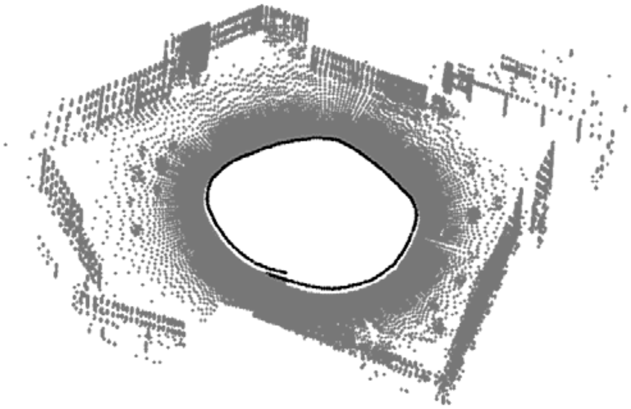


Fig. 11. Outdoor result of motion estimation and 3D reconstruction (Black: estimated motion, Gray: reconstructed 3D structure)

TABLE II
OUTDOOR MOTION ESTIMATION ERROR

Frame	Length	Error	Error rate
300	110m	3.68m	3.3%

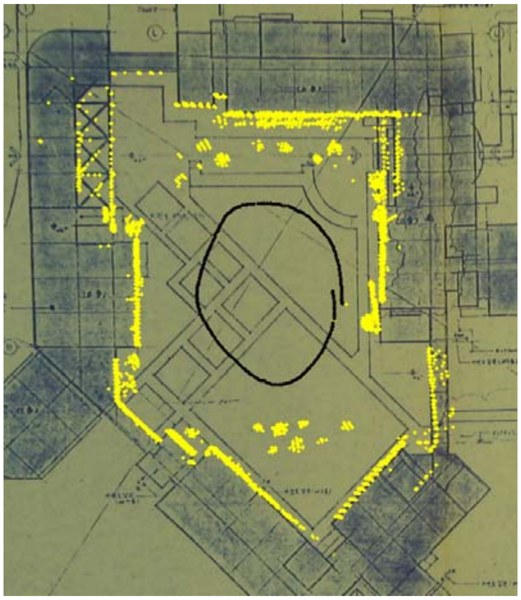


Fig. 12. Overlapping the result with the floor plan of the building (Yellow: reconstructed walls, Black: estimated motion)

connected to itself in reverse order, a closed-loop sequence, a sequence with known true motion, can be obtained. We estimate motion of the generated sequence while adjusting the length of it. The result is shown in Fig. 14. The error is nearly proportional to the length of the sequence, and the error rate is below 1%.

D. *Computation Time*

We check the computation time needed for our proposed method. The computer used for this paper has a 2 GHz CPU. The total process consists of three major parts: matching, computation of initial solution, and nonlinear optimization. The average computation time needed for each process is given in Table III. More than one second is needed to estimate motion of a frame. However, this time can be reduced; e.g. the time for matching decreases if the amount of data to be matched is smaller, and the time for the initial solution decreases if the number of iterations in the RANSAC [18] process is lower. Table IV shows the reduced computation time using 1/2 of the laser and feature points and 1/10 of the number of iteration. The time for initial solution considerably decreases, but the time for matching doesn’t decrease because of the KLT implementation. If the time for matching decreases by using other method, this system can be adapted to real time applications.

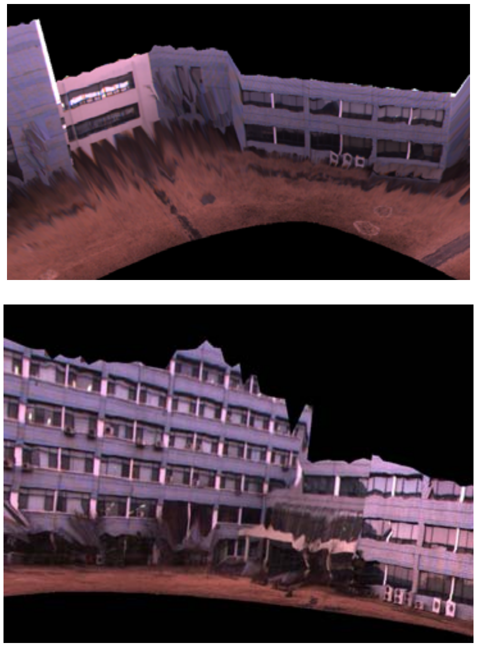


Fig. 13. Parts of the result with texture mapping

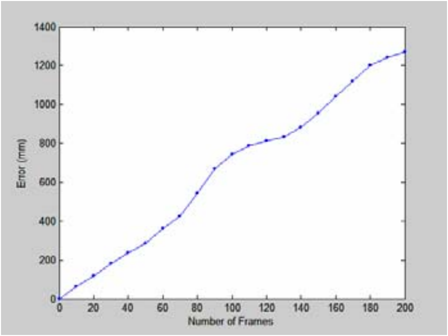


Fig. 14. Closed-loop error using real data

TABLE III
AVERAGE COMPUTATION TIME

Matching	Initial Solution	Optimization
0.709 sec	0.947 sec	0.003 sec

TABLE IV
REDUCED COMPUTATION TIME

Matching	Initial Solution	Optimization
0.688 sec	0.076 sec	0.002 sec

V. Conclusion

In this paper, we present a new sensor system for motion estimation and 3D reconstruction. Using a camera-based motion estimation method, we cannot compute accurate 3D structures due to the limitation of image resolution without any constraints. Using a 2D laser range finder, we cannot compute 3D motion without a tilting module. We combine a camera and a 2D laser sensor to complement each other. For this system, we propose a new algorithm that uses scan data as 3D points of the conventional 3D-to-2D method. This algorithm is structured to avoid the degenerate case of the conventional 3-point algorithm. Our algorithm proved to be more reliable for the proposed system than other algorithms. Our sensor system and algorithm provided accurate results in both indoor and outdoor experiments.

The proposed system fusing two different types of sensors can be a practical solution for motion estimation and 3D reconstruction.

Acknowledgment

This research is supported by the National Research Laboratory (NRL) program (No. M1-0302-00-0064) of the Ministry of Science and Technology (MOST), and the Agency for Defense Development (ADD).

References

[1] D. Nistér, "An Effective Solution to the Five-Point Relative Pose Problem", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004.

[2] D. Nistér, O. Naroditsky, and J. Bergen, "Visual Odometry", in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.

[3] A. J. Davison, "Real-Time Simultaneous Localization and Mapping with a Single Camera", in *Proceedings of the IEEE International Conference on Computer Vision*, 2003.

[4] A. Nüchter el al., "6D SLAM with an Application in Autonomous Mine Mapping", in *Proceedings of the IEEE International Conference on Robotics & Automation*, 2004.

[5] A. Banno and K. Ikeuchi, "Shape Recovery of 3D Data Obtained from a Moving Range Sensor by using Image Sequences", in *Proceedings of the IEEE International Conference on Computer Vision*, 2005.

[6] Q. Zhang and R. Pless, "Extrinsic Calibration of a Camera and Laser Range Finder (improves camera calibration)", in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2004.

[7] Z. Zhang, "Flexible Camera Calibration by Viewing a Plane from Unknown Orientations", in *Proceedings of the IEEE International Conference on Computer Vision*, 1999.

[8] J. S. Kim, P. Gurdjos, and I. S. Kweon, "Geometric and Algebraic Constraints of Projected Concentric Circles and Their Applications to Camera Calibration", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005.

[9] R. Haralick et al., "Review and Analysis of Solutions of the Three Point Perspective Pose Estimation Problem", *International Journal of Computer Vision*, 1994.

[10] D. Nistér, "A Minimal Solution to the Generalised 3-Point Pose Problem", in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.

[11] D. Ortín, J. Neira, and J. M. M. Moltiel, "Relocation using Laser and Vision", in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2004.

[12] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", *International Journal of Computer Vision*, 2004.

[13] J. Shi and C. Tomasi, "Good Features to Track" in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1994.

[14] C. J. Harris and M. Stephens, "A combined corner and edge detector", in *Proceedings of the Alvey Vision Conference*, 1988.

[15] S. Kim, K. J. Yoon, and I. S. Kweon, "Object Recognition using Generalized Robust Invariant Feature and Gestalt Law of Proximity and Similarity", *IEEE Workshop on Perceptual Organization in Computer Vision (in CVPR'06)*, 2006.

[16] P. Gurdjos, J. S. Kim, and I. S. Kweon, "Euclidean Structure from Confocal Conics: Theory and Application to Camera Calibration", in *Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006.

[17] P. J. Besl and N. D. McKay, "A Method for Registration of 3-D Shapes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992.

[18] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography", *Comm. of the ACM*, 1981.