CALOS : Camera And Laser for Odometry Sensing

Yunsu BOK, Youngbae HWANG and In-So KWEON

RCV Lab., School of Electrical Engineering and Computer Science, KAIST.
{ysbok, unicorn}@rcv.kaist.ac.kr, iskweon@kaist.ac.kr

Abstract The CCD camera and the 2D laser range finder are widely used for motion estimation and 3D reconstruction. With their own strength and weakness, 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 non-linear 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.

1 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 whole 3D scene. But 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. But we cannot obtain the depth information of the whole 3D structure from scan data without additional device such as tilting module.

Some methods have been proposed to estimate motion in 3-D space using CCD cameras. 5-point algorithm [1] and 3-point algorithm [2][8][9] estimate the initial motion using some point correspondences and minimize the re-projection error. Probabilistic approaches based on the extended Kalman filter (EKF) or particle filter provide good motion estimation results [3]. But the 3D reconstruction results are not so 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 of 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 some constraints from the camera motion and the laser data. Both sensors are also used in [10]. In indoor environment, the motion of a robot is estimated using a 2D laser sensor. Then the image is transformed based on this result to make feature matching easier.

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 single camera based approach 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 cameras and a 2D laser sensor. 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.

2 A New Sensor System : CALOS

A new sensor system called CALOS(Camera And Laser for Odometry Sensing) consists of cameras and a 2D laser sensor.

2.1 SingleCALOS : 1 Camera

SingleCALOS, the combination of a camera and a laser sensor, is the simplest type of the CALOS sensor system. An example of SingleCALOS is shown in Fig.1. Only two sensors are needed for this system. But a degenerate case exists because of narrow field of view. (The degenerate case will be discussed later.)

2.2 MultiCALOS : 2 or More Cameras

MultiCALOS consists of 2 or more cameras and a laser sensor, as shown in Fig.2. The field of view which depends on the arrangement of the cameras can be expanded. Some problems also exist, such as synchronization and computational complexity. But a wider field of view reduces the errors in motion estimation and 3D reconstruction.

Fig. 1: SingleCALOS

Fig. 2: MultiCALOS

3 Motion Estimation & 3D Reconstruction

3.1 Motion Estimation : Basic Idea

The basic concept of motion estimation for CALOS is 3D-to-2D matching. It means that we minimize the projection error of 3D points on image plane. The most important feature of CALOS is that the laser data are used as the 3D points. The 3D points obtained from single or stereo camera system are inaccurate due to the limitation of image resolution. On the contrary, the laser points have very small error which is independent of the measured range. (For example, the sensor used in this paper has ±10mm error at the distance of 5m, 10m, and 20m) This configuration is equal to that of conventional 3-point algorithm, but the 3D points are very accurate.

3.2 Extrinsic Calibration

To use images and range data simultaneously, it is necessary to transform data into a common coordinate. 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 relative pose between the sensors is assumed to be fixed. The camera is calibrated first using a pattern plane [7]. Then the position of the plane in camera coordinate is computed. While the camera captures the images of the plane, the laser sensor also scans the same plane. The relative pose between the sensors is estimated using the fact that some points of the scanned data are on the pattern plane.

3.3 Correspondence Search

To find the location of the current scan data on the next image, we project the current scan data on the current image and find the corresponding point on the next image, as shown in Fig.3. Some methods for image matching have been proposed. Template matching, KLT [12], descriptor matching of features such as SIFT [11] and G-RIF [14] are good examples. In this paper, we implement an SAD based method.

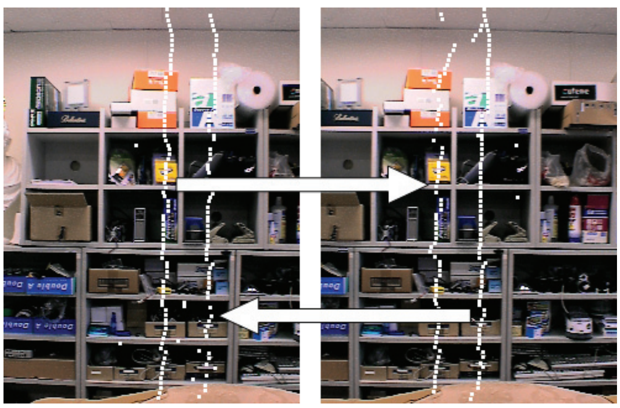


Fig. 3: Example of correspondence search

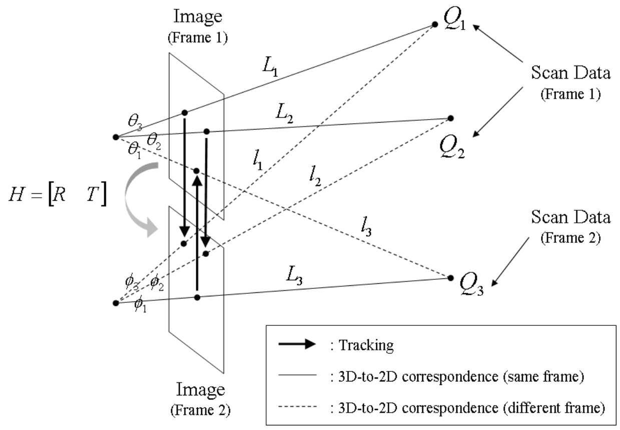


Fig. 4: Initial motion for SingleCALOS

3.4 Initial Solution for SingleCALOS

A degenerate case of conventional 3-point algorithm exists because we use a 2D laser sensor. If it scans a plane, the scanned data are on a line. To avoid this case, 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.4, 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 them, and then $L_1 \sim L_3$ are computed. The angles between the camera rays, $\theta_1 \sim \theta_3$ and $\phi_1 \sim \phi_3$, are also computed if we find the correspondences of projected laser points. (Solid lines and dotted lines represent the 3D-to-2D correspondences; the former is the correspondence between the images and scan data of same frame while the latter is that of different frame.) Unknown lengths are $l_1 \sim l_3$. Applying the second law of cosine, Eq.(1)~(3) are derived from Fig.4.

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

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

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

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

3.5 Non-linear Optimization

We refine the initial solution by the non-linear optimization of a cost function:

$$\sum \frac{(q'^T Eq)^2}{(e_1q)^2 + (e_2q)^2} + \sum \left(p - \frac{[R \mid T]P}{[r_3 \mid t_3]P} \right)^2 \quad (4)$$

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

where q and q' are the corresponding feature points on the images. P and p are the laser point and its correspondence on the other image. The first term of Eq.(4) means the distance between the epipolar line and the corresponding point on the image. The second term means the projection error on 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 [13].

3.6 Initial Solution for MultiCALOS

A wider field of view of the MultiCALOS system reduces the degenerate cases and we may use the conventional 3-point algorithm [8]. But the MultiCALOS uses more cameras than one. It means that the rays don't meet at a point. For the case of arbitrary rays, the generalized 3-point algorithm [9] is used, as shown in Fig.5. If Eq.(6) is the transformation matrix from the laser coordinate to the camera coordinate, the origin p and direction d of the ray are computed by Eq.(7). (u, v) is the normalized coordinate of corresponding point of the laser point. MultiCALOS also needs the non-linear optimization based on the same cost function.

$$H = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \quad (6)$$

$$p = -R^T T \qquad d = R^T [u \quad v \quad 1]^T \quad (7)$$

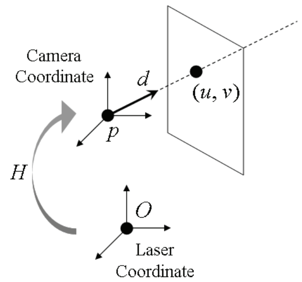


Fig. 5: Camera ray information for MultiCALOS

3.7 3D Reconstruction

The 3D reconstruction problem is solved easily with the proposed CALOS sensor system. If the motion of the system is estimated, the scan data are transformed into a common coordinate. In Eq.(8), 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 laser coordinate system of the first frame. And the texture mapping is easy because all of the scan data are projected onto the images.

$$l_{1n} = \begin{bmatrix} R_1 & T_1 \\ 0 & 1 \end{bmatrix}^{-1} \cdots \begin{bmatrix} R_{n-1} & T_{n-1} \\ 0 & 1 \end{bmatrix}^{-1} L_n = \left(\prod_{i=1}^{n-1} \begin{bmatrix} R_i & T_i \\ 0 & 1 \end{bmatrix}^{-1} \right) L_n \qquad (8)$$

The feature points on the images also can be added to the 3D points. If the matching is correct, triangulation generates the 3D information of wider range than the laser data. But the points generated by the feature points have large uncertainty in the heading direction of cameras. If it is impossible to reduce the uncertainty using many images, only the laser points should be included to the 3D reconstruction.

4 Experimental Results

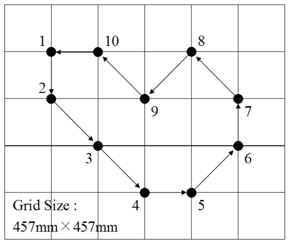
In this paper, we use SONY DFW-V500 cameras and SICK LMS200 laser sensor. The image size is 320×240.

4.1 SingleCALOS : Motion Estimation

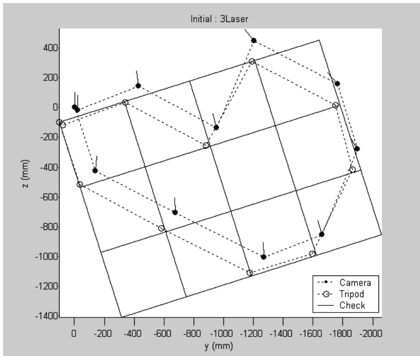
The SingleCALOS is fixed on a tripod, as shown in Fig.1, for indoor experiments. The image and scan data are obtained at every node of a grid pattern on the floor (see Fig.6(a)). Fig.6(b) shows the result of motion estimation. The total length of motion is 5.7m, and the closed-loop error is about 50mm (less than 1%). The maximum error of tripod position at all nodes is less than 50mm.

4.2 SingleCALOS : 3D Reconstruction

The SingleCALOS is attached to a vehicle for outdoor experiment. The reconstruction result is shown in Fig.7. We transform the laser points into a common coordinate using Eq.(8), and perform texture mapping. The result shows the structure of a building, ground, and trees. Using CALOS, realistic 3D structure is reconstructed easily by just accumulating scan data.



(a) True motion of tripod



(b) Result (dot: camera, circle: tripod)

Fig. 6: Motion estimation (SingleCALOS, indoor)

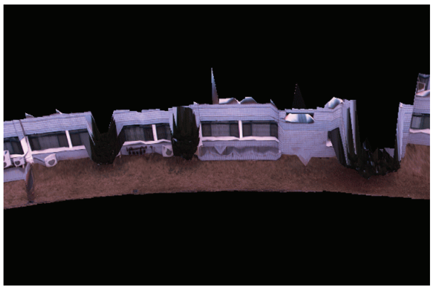
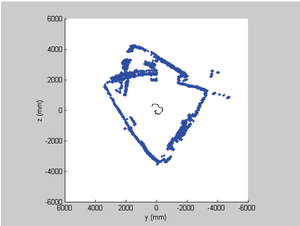


Fig. 7: 3D reconstruction (SingleCALOS, outdoor)

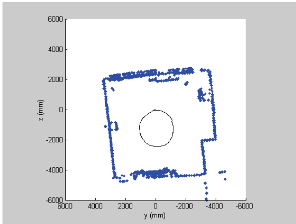
4.3 SingleCALOS vs. MultiCALOS

To compare two systems, SingleCALOS and MultiCALOS, we perform the 3D reconstruction experiment in indoor environment. Fig.8 shows the cross section of reconstructed 3D model. The result by MultiCALOS is accurate while that by SingleCALOS shows some error accumulation problem, as shown in Fig.8(a) and Fig.8(b). The circular shape at the center of Fig.8 is the motion of CALOS. The result by MultiCALOS is better than that by SingleCALOS. It is because the additional cameras of MultiCALOS face the floor and ceiling, and help reducing the motion ambiguity.

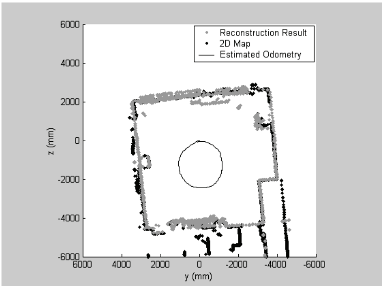
The performance of CALOS depends on the number of cameras. Many cameras provide the wide field of view which makes the solution accurate and stable.



(a) SingleCALOS



(b) MultiCALOS



(c) Overlapping with 2D map

Fig. 8: 3D reconstruction (indoor)

4.4 MultiCALOS : 3D Reconstruction

Fig.9 shows the sensor trajectory and the result of 3D reconstruction in outdoor environments. To verify the accuracy of the result, we compare the reconstruction result to the floor plan of the building. Fig.10 shows that the result overlaps the floor plan well.

5 Conclusion

In this paper, we present a new sensor system for motion estimation and 3D reconstruction. We combine cameras and a 2D laser sensor to support each other. For this system, we propose a new algorithm that uses scan data as 3D points of conventional 3D-to-2D method. The proposed sensor system and algorithm provide accurate results in both indoor and outdoor experiments.

The performance of the sensor system depends on the number of the cameras. We have two choices, SingleCALOS for fast computation and MultiCALOS for accurate result. The proposed system fusing two different types of sensors can be a practical solution for motion estimation and 3D reconstruction.

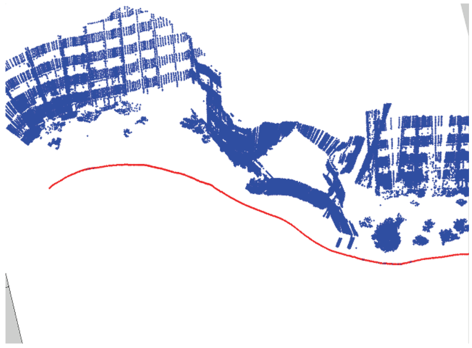
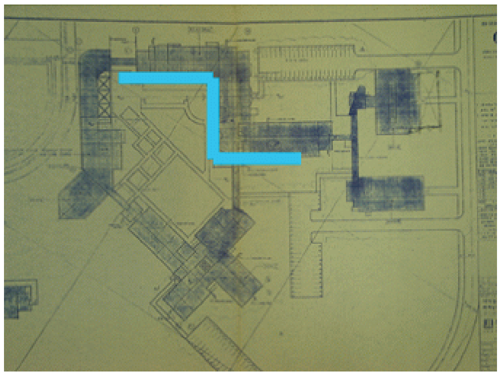
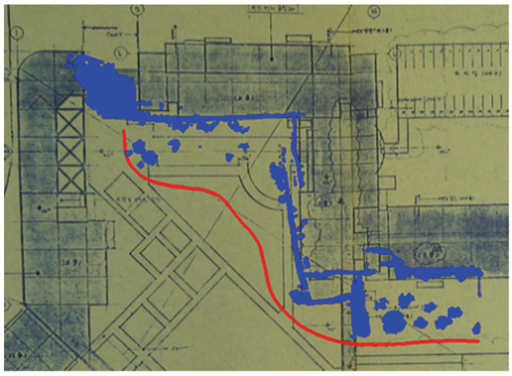


Fig. 9: 3D reconstruction (MultiCALOS, outdoor)



(a) Floor plan of a building
(Bright part is reconstructed)



(b) Overlapping with result

Fig. 10: Validation of result in Fig.9

Acknowledgment This research is supported by National Research Laboratory (NRL) program (No. M1-0302-00-0064) of Ministry of Science and Technology (MOST), and 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, 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, K. Ikeuchi, "Shape Reconvery 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, 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] R. Haralick et al., "Review and Analysis of Solutions of the Three Point Perspective Pose Estimation Problem", International Journal of Computer Vision, 1994.
[9] 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.
[10] D. Ortín, J. Neira, J.M.M. Moltiel, "Relocation using Laser and Vision", in Proceedings of the IEEE International Conference on Robotics and Automation, 2004.
[11] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, 2004.
[12] J. Shi, C. Tomasi, "Good Features to Track" in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994.
[13] C.J. Harris and M. Stephens, "A combined corner and edge detector", in Proceedings of the Alvey Vision Conference, 1988.
[14] S. Kim, K.J. Yoon, 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.

Yunsu BOK: received the B.S. degree and the M.S. degree in electrical engineering from Korea Advanced Institute of Science and Technology, Korea, in 2004 and 2006, respectively. He is a Ph.D. student of the school of electrical engineering and computer science in KAIST. His research interest includes motion/odometry estimation, sensor fusion, and camera calibration.

Youngbae HWANG: received the B.S. degree and the M.S. degree in electrical engineering from Korea Advanced Institute of Science and Technology, Korea, in 2001 and 2003, respectively. He is a Ph.D. student of the school of electrical engineering and computer science in KAIST. His research interest includes noise modeling, image segmentation, motion/odometry estimation.

In-So KWEON: received the B.S. degree and the M.S. degree in machine design from Seoul National University, Korea, in 1981 and 1983, respectively. He received the Ph.D. degree in robotics from Carnegie Mellon University, USA, in 1990. He is a professor of the school of electrical engineering and computer science in KAIST from 1992. His research interest includes computer vision, 3D vision, camera calibration, object recognition, object classification, SLAM. He is a member of the IEEE, and Korea Robot Society (KRS).