Stereo Vision based Motion Adjustment of 2D Laser Scan Matching

Sung-In Choi, Lei Zhang
Electrical Engineering and Computer Science
Kyungpook National University
Daegu, Republic of Korea
ellim5th@naver.com, tinysand0527@gmail.com

Soon-Yong Park
Computer Science and Engineering
Kyungpook National University
Daegu, Republic of Korea
sypark@knu.ac.kr

***Abstract* – In this paper, we considered the problem of SLAM with a combination of laser range finder and stereo vision camera in an indoor environment. Iterative Closest Point(ICP) is one of the most popular algorithms for robot's motion estimation. However, because of its fundamental limitation, it is hard to maintain accurate pose of the robot if there are no distinguishable features such like corners. To overcome this problem, we employ a stereo vision camera as a motion adjustment unit. We split matching and adjustment parts into two separate threads. Using the parallel processing scheme, the motion can be estimated in real-time. Experimental results illustrate a great improvement comparing with a variant of ICP and proposed method.**

Keywords – stereo vision, ICP, SLAM, map building

I. INTRODUCTION

A laser range finder has been widely used for Simultaneous Localization and Mapping (SLAM) in the robotics field because of its accurate precision and fast sensing ability. By matching two consecutive range scans captured from the laser range finder, the relative motion can be computed directly. A number of range scan matching techniques have been introduced and successfully adopted to estimate the relative pose of mobile robots[1][2][3][4]. Among them, the ICP algorithm which is firstly introduced by Besl and McKay[5] has been widely used for robot localization since it provides quite reasonable results with a simple formulation and a motion model. But, because of its fundamental limitation, it is hard to maintain accurate pose if there are no distinguishable features such like corners. Fig. 1 shows an example, in which the mobile robot failed to estimate its position when the ICP algorithm was applied as a motion estimator in the long corridor region. Moreover, an accumulation error can be shown in terms of global positioning and it has been regarded as a conventional problem in pairwise matching scheme. To overcome this problem, in this paper, we additionally employ a stereo vision sensor to enhance the accuracy of range scan matching.

The basic concept of our proposed method is to check the registration error of range scan matching using a vision based measurement and to adjust the incorrectly estimated pose using a stereo vision based motion estimation technique. According to this basic concept, our proposed system consists of two parts: incremental motion estimation using a laser range finder and key-frame adjustment using a stereo vision sensor.

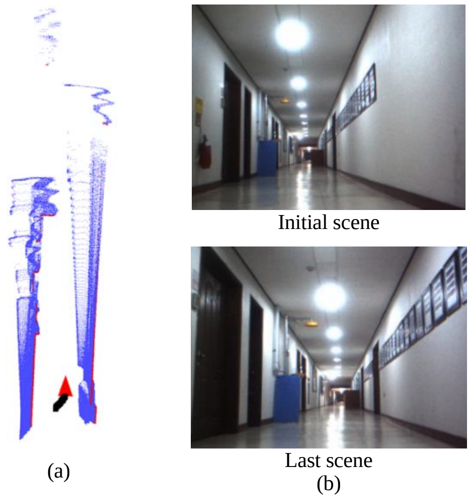


Figure 1. Limitation of the ICP algorithm in a long corridor. (a) top view of miss-aligned laser range scans (b) snapshots of initial and last frames

Most of earlier works utilize the adjustment technique at every distinctive frame as soon as the motion estimation is finished in a same thread. In general, the key-frame adjustment technique requires too much processing time, thus it used to fail to maintain the robot's pose because the motion estimation part will be in an idle state during the adjustment in spite of continuous motion of the robot. In this case, it needs to slow down moving speed of the robot to maintain correct pose. To solve this problem, we propose to split matching and adjustment parts into two separate tasks, processed in parallel threads on a dual or quad core CPU: one thread deals with the task of the incremental motion estimation of the robot, while the other verifies the accuracy of estimated pose between reference frame and current frame and refines the pose when error is exceed above a pre-defined threshold value.

II. HYBRID LASER-STEREO SLAM SYSTEM

A. System Overview

Fig. 2 is a schematic of our proposed two-stage SLAM system that combines laser range scan based localization and stereo vision based motion refinement. As shown in Fig. 2, our proposed system consists of four threads: two threads are assigned for data acquisition of the laser range finder and the stereo vision sensor and another two threads are assigned for matching and adjustment tasks, respectively. The data

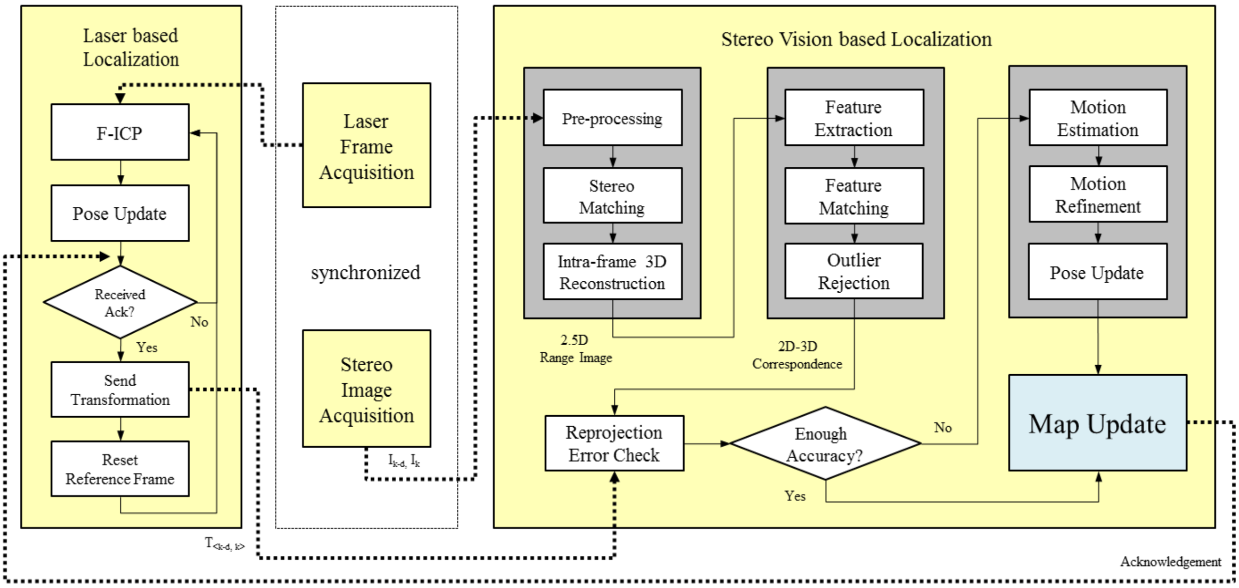


Figure 2. Overview of the proposed hybrid laser-stereo SLAM framework.

acquisition threads are completely synchronized. When the system initializes, the initial pose of the robot $r_0=[X_0, Y_0, \Phi_0]^T$ at time 0 is set as $[0, 0, 0]^T$ and the first frame of range scan is set as a reference frame. Then pairwise motion estimation by laser scan matching is started.

After several frames, the adjustment thread sends a signal to the laser part to receive an accumulated pose data from the reference frame to the current frame with stereo images which are captured by the stereo camera at the reference frame and the current frame, respectively. Then the laser part resets the current frame as a new reference frame and starts again the pairwise motion estimation process immediately. At the same time, the stereo vision part receives data for adjustment and also starts the verification process. If the accuracy of transformation is good enough, then map updating is performed directly. Otherwise, the stereo vision based motion estimation process is additionally followed. At the end of stereo vision based motion estimation, an accuracy check process is performed one more time. If the result of laser part shows better accuracy then vision estimation, then it is used for map updating as an optimal solution. But if it's not, the vision based estimation is applied for map updating. At last, the adjustment thread sends a signal to the laser part and iterates again all of the procedures.

B. *Extrinsic Calibration*

To fuse two different sensors for localization in a common coordinate system, it is necessary to find a relative pose between the laser range finder and the stereo vision sensor. Since all 3D points of the stereo sensor are computed in the right camera coordinate system, we also set the right camera coordinate as a reference coordinate of the stereo sensor. A method for calibration between each sensor is proposed in [6]. The relative pose between each sensor is calculated by constraining the fact that the pose of a pattern plane can be calculated in the stereo sensor coordinate by camera calibration and the scanned laser data contains the laser points which lie on the pattern plane. For more detailed information about non-linear optimization, please refer to [6].

C. *Incremental Motion Estimation using Laser Range Finder*

In this paper, we utilize a variant of ICP algorithm for incremental motion estimation of our mobile robot. During navigation, we can easily face miss-aligning situation if we apply conventional ICP algorithm for motion estimation due to outliers, occlusions and sensor noise. Since the consecutive laser scans only overlap partially, the criterion of the closest point can result many false correspondences from the non-overlapping area. In this paper, we use the Fractional ICP algorithm to reject false correspondences effectively. Only the correspondences which minimize the Fractional RMSD (Root Mean Square Distance) are used for Least Squares minimization.

D. *Stereo Vision based Key-frame Adjustment*

1) *Intra-frame 3D Reconstruction*

Our stereo vision based motion adjustment method utilizes a 2.5D range image which contains a number of 3D points and their associated 2D feature points in a same structure form. The 2.5D range image can be generated by stereo matching algorithm. The entire procedure of generating the 2.5D range image consists of following three steps:

- Rectification aligns the epipolar lines in the same image rows.
- Smoothing removes high-frequency component from the rectified images. We applied the Gaussian filter at every frame.
- Stereo matching & 3D reconstruction: to establish correspondence between stereo images, the Sum of Absolute Differences correlation is used. In addition, uniqueness and texture validation is also applied as a matching constraint.

2) *Motion Verification*

The main idea of the verification is re-projection error check between reference and current frames with the transformation T^L accumulated by laser scan matching. Suppose there are two sets of 2.5D range images with an interval d at current time t: $I^R = I_{t-d} = \{x_{t-d}, X_{t-d}\}$ obtained from the reference frame and $I^C = I_t = \{x_t, X_t\}$ obtained from the current frame. Here, x denotes the 2D image features and X means the corresponding 3D points of those image features. Each set of 3D points X is defined in the camera coordinate system. Also, matching relation between I^R and I^C is already established by some kind of image matching technique and indexed from 0 to N. Then the quality of estimated transformation $T^L_{<t-d,t>}$ from reference frame to current frame can be determined by the following cost function of re-projection error:

$$f_{reproj}\ (P, T^L_{<t-d,t>}, I^R, I^C) = \sum_i^N [(x^i_{t-d} - PT^L_{<t-d,t>}{}^{-1} X^i_t)^2 + (x^i_t - PT^L_{<t-d,t>} X^i_{t-d})^2] \quad (1)$$

where P is the 3x4 camera projection matrix. In general, the re-projection error is smaller than 0.5 pixel then we can consider the laser scan matching was successful. Due to the previous intra-frame 3D reconstruction stage, a number of 3D points with its corresponding image points are already acquired. The remaining for verification is to find a test set (e.g. 2D-2D and 3D-3D) using image matching technique between two 2.5D range images.

The key-frame image matching will be performed on unknown camera ego-motion. Thus it is more suitable to adopt a viewpoint-invariant matching technique such as SIFT[7], SURF[8] for robustness. In this paper, we employ the SURF algorithm since it gives not only high speed matching performance but also reliable accuracy. We applied Schulz's GPU implementation in order to maximize the computation time[9]. In addition, for effective outlier rejection, we also applied the epipolar constraint with a RANSAC approach using the 8-point algorithm to check the matching consistency[10]. The RANSAC is performed on maximum 1.0 pixel error with 99% confidence level. Similar to the 2D-2D matching case, the established 3D-3D matching pairs also may have a number of outliers due to sensing limitation or poor 3D reconstruction. In order to solve this problem, a simple rigidity constraint based on the RANSAC algorithm is additionally applied[11].

3) *Motion Recovery and Refinement*

After finishing the motion verification stage, we can judge the quality of the result of laser scan matching. If the quality is sufficiently accurate, then we directly use the result of laser scan matching $T^L_{<t-d,t>}$ to update the map and robot's pose at time t. Otherwise, assuming the failure of laser scan matching, we try to find another solution using stereo vision based motion estimation. Thanks to the previous verification stage, we already know 3D-3D correspondences between two key-frames by applying the SURF matching to the 2.5D range images. If a sufficient number of 3D matching pairs $\{X_t, X_{t-d}\}$ are given, a transformation matrix $T^S_{<t-d,t>} = [R^S_{<t-d,t>} \mid t^S_{<t-d,t>}]$ which minimizes the robot's pose error e can be calculated by following energy function:

$$e = \sum_i^N \left\| X^i_{t-d} - (R^s_{<t-d,t>} X^i_t + t^s_{<t-d,t>}) \right\| \quad (2)$$

Finding initial transformation matrix $T^S_{<t-d,t>}$ using Eq. (2) can be achieved by a Least-Squares Fitting algorithm[12]. The refinement process is done using a nonlinear optimization algorithm such as Levenberg-Marquardt method[13]. For the non-linear optimization, Eq. (1) was utilized again as a LM-minimization cost function.

4) *Key-frame Selection*

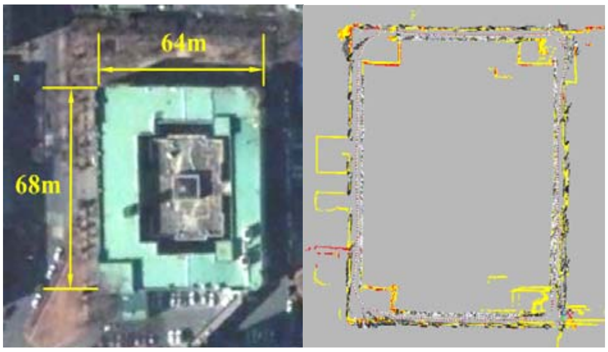
Most of previous vision based SLAM approaches select the key-frame when tracking feature points have enough disparity range or there are insufficient tracking points between reference frame and current frame[14][15]. Because of this selecting condition, it is necessary to check the count of tracking points at every image frame and it becomes a time wasting. Compared to the earlier works, our proposed system uses a simple but convincing method. Right after the map updating is finished, our system immediately selects the recently retrieved laser-stereo data as a new key-frame.

III. EXPERIMENTAL RESULTS

To evaluate the performance of our proposed system, we captured the Engineering Building No. 5 at Kyungpook National University(KNU) in Daegu, Korea, which is one of the large-scale rectangular buildings in KNU. There is no ground truth about the size of the building. Thus we approximately measured the size of the building using Google Earth to check localization error. Fig. 4-(a) shows an orthogonal view of the building with measured size. The mobile robot was navigated by remote control to acquire closed-loop sequences. The moving speed of the robot was 0.5m/s. The length of the path is about 184 meters and totally 17970 synchronized laser-stereo frames are captured during the navigation for the experiment. Among them, 3872 frames are selected as key-frames for motion adjustment and map update. All the range frames were scanned by Hokuyo UTM-30LX laser range finder. It has 270° field of view(FOV) and 0.25° angular resolution. As a stereo sensor unit, Bumblebee 2 manufactured by Point Gray Research is used. The image size was reduced to 320x240 pixels. As shown in Fig. 3, both of sensors are mounted on the mobile robot to look forward. All the processing was done offline on a 2.4GHz Intel i7 multi-core processor PC equipped with 4.0GB of RAM. To support GPU computation, NVIDIA GTX260 is also employed.

Fig. 5 shows a comparison of original ICP(cyan), Fractional-ICP(blue), stereo visual odometry(green) and proposed method(red). The start point is denoted by a black cross mark. As you can see in Fig 5., our proposed method gives the most reliable accuracy compare to other methods. The Fractional-ICP algorithm also draws a closed-loop but the scale of drawn trajectory is not realistic because of its accumulation error. Fig. 4-(b) shows the orthogonal view of the reconstructed 3D map. The map built by the proposed method is consistent with the real structure of the building. The average computation time when stereo based adjustment is occurred is given in Table 1.

Figure 3. A mobile robot with sensors used in the experiments



(a) (b)

Figure 4. Experimental site: (a) top view of Engineering Building No. 5 at KNU (b) top view of a reconstructed 3D map

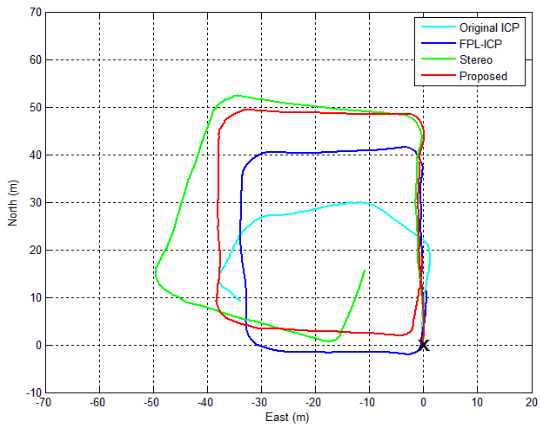


Figure 5. Accuracy comparison of four methods: original ICP(cyan), Fractional-ICP(blue), stereo visual odometry(green) and proposed method(red)

TABLE I. PROCESSING TIME ANALYSIS

Step	Time / Frame (ms)
2.5D range image genetaion	31.21
Motion verification	51.17
Motion recovery & refinement	13.42
Map updating	2.41
Total	**98.21**

IV. CONCLUSION

In this paper, we considered the problem of SLAM with a small unmanned ground vehicle in an indoor environment using a combination of laser and stereo sensors. To overcome the fundamental limitation of laser scan matching for motion estimation, we have presented a set of stereo vision based adjustment technique. By integrating with stereo vision sensor measurements during the robot navigation, the accuracy of robot's pose has been dramatically improved. In the future, the proposed localization method will be mounted on the small unmanned ground vehicle to test autonomous navigation in real-time.

ACKNOWLEDGMENT

This research has been supported by funding from FM electronics Inc. and the Agency for Defense Development in Korea.

REFERENCES

[1] A. Censi, "An ICP variant using a point-to-line metric," in Proc. of IEEE International Conference on Robotics and Automation, pp. 19-25, 2008.

[2] J. Minguez, ,F. Lamiraux and L. Montesano, "Metric-based scan matching algorithms for mobile robot displacement estimation, " in Proc. of IEEE International Conference on Robotics and Automation, pp. 3557-3563, 2005.

[3] S. Hong, H. Ko and J. Kim, "VICP: velocity updating iterative closest point algorithm", in Proc. of IEEE International Conference on Robotics and Automation, pp. 1893-1898, 2010.

[4] D. Holz and S. Behnke, "Sancta simplicitas – on the efficiency and achievable results of SLAM using ICP-based Incremental Registration, " in Proc. of IEEE International Conference on Robotics and Automation, pp. 1380-1387, 2010.

[5] P. Besl and N. McKay, "A method for registration of 3-D shapes," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pp. 239-256, 1992.

[6] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," in Proc. of the IEEE/RSJ International Conference on Inteligent Robots and Systems, 2004.

[7] D. G. Lowe, "Distinctive Image Features from Scale-invariant Keypoints," International Journal of Computer Vision, Vol. 60, No. 2, pp. 91-110, 2004.

[8] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346-359, 2008.

[9] A. Schulz, F. Jung, S. Hartte, D. Trick, C. Wojek, K. Schindler, "CUDA SURF - A real-time implementation for SURF," in http://www.d2.mpi-inf.mpg.de/surf?q=surf, 2011.

[10] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision," 2nd Ed, Cambridge University Press, 2004.

[11] A. Howard, "Real-time stereo visual odometry for autonomous ground vehicles," in Proc. International Conference on Intelligent Robots and Systems, pp. 3946-3952, 2008.

[12] K. S. Arun, T. S. Huang, S. D. Blostein, "Least-Squares Fitting of Two 3-D Point Sets", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 9, No. 5, pp. 698-700, 1987.

[13] K. Levenberg, "A method for the solution of certain non-linear problems in least square," The Quarterly of Applied Mathmatics, pp. 164-168, 1944.

[14] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd, "Real time localization and 3D reconstruction," in Proc. of IEEE International Conference on Computer Vision and Pattern Recognition, 2006.

[15] Z. Zhu, T. Oskiper, O. Naroditsky, S. Samarasekera, H. S. Sawhney, and R. Kumar, "Ten-fold Improvement in Visual Odometry Using Landmark Matching," in Proc. International Conference on Computer Vision, 2007.