TEST Engineering & Management

November-December 2019
ISSN: 0193-4120 Page No. 949 - 955

SLAM-based Laser Scanning for 3D Reconstruction

Jie Gao[1], Hao Fan[2], Jun Yu Dong *[3], Shu Zhang [4], Lin Qi[5], Hui Yu[6]
[1,2,4,5]*Ocean University of China, College of Information Science and Engineering, Qingdao, China*
gaojie9186@stu.ouc.edu.cn, fanhao@ stu.ouc.edu.cn, zhangshu@ouc.edu.cn, qilin@ouc.edu.cn
[3]*Ocean University of China, College of Information Science and Engineering, Qingdao, China*
**Corresponding author:junyu.dong@outlook.com*
[6]*University of Portsmouth, School of Creative Technologies, Portsmouth, United Kingdom*
hui.yu@port.ac.uk

Article Info
Volume 81
Page Number: 949- 955
Publication Issue:
November-December 2019

Article History
Article Received: *3 January 2019*
Revised: *25 March 2019*
Accepted: *28 July 2019*
Publication: *25 November 2019*

Abstract

Conventional laser scanning typically uses sensors to estimate camera pose and complete 3D reconstruction. For some cases, linear scanning is used to achieve point cloud stitching for less computational costs, however, seriously affects the practicality and flexibility. Moreover, it can only achieve semi-dense reconstruction. This paper proposes a laser scanning based 3D reconstruction combining with Simultaneous Localization and Mapping (SLAM) to tackle aforementioned challenges, SLAM provides pose estimations and accurate point cloud stitching outputs. The pipeline of the proposed method is twofold:(a) a three-dimensional semi-dense point cloud reconstruction of a target object using SLAM-based laser point cloud stitching; (b) a robust 3D-3D alignment scheme, which makes the point cloud obtained by SLAM can be merged with the laser scanning result. The results back-projected onto the original image have very limited deviations after point cloud fusion. The experiments demonstrate that our method can provide a high accuracy rate and produce a high-quality 3D surface with fine geometric details.

***Keywords**:* 3D reconstruction, Laser scanning, Simultaneous Localization and Mapping (SLAM),Equal-scale mapping

1. INTRODUCTION

High-quality dynamic three-dimensional (3D) reconstruction has become an active research topic with multiple practical applications[1][2], such as medical image processing, obstacle avoidance for unmanned vehicles and augmented reality (AR). There are many methods for dynamic 3D modelling, such as structure from motion (SFM) [3], multi-view stereo [4], laser scanning (LS) [5], and simultaneous positioning and mapping (SLAM) among others. These methods have produced amazing results in most cases. For example, traditional laser scanner can reconstruct semi-dense high-precision point cloud. However, it usually relies on dedicated sensors to estimate camera pose and complete 3D reconstruction, which is not flexible and sometimes requires strict constraints and expensive setup. In recent years, the research of SLAM has made significant progress and is considered to be one of the most advanced real-time 3Dreconstruction methods. Among them, Stereo SLAM can perform pose estimation with absolute scale. One

November-December 2019
ISSN: 0193-4120 Page No. 949 - 955

of their downsides is that the sparse SLAM method [6][7] tracks a set of image feature points to solve camera motion and build sparse 3D maps from these tracking points. Therefore, it cannot recover a complete surface of the object specifically and accurately; Although the dense SLAM method [8][9][10]can produce a dense depth, it has high computational requirements and demands the most advanced GPU to run in real time. In this paper, we propose a laser scanning based 3d reconstruction combining with SLAM. It can obtain a local high-precision object structure by using only a binocular camera and a laser projector.

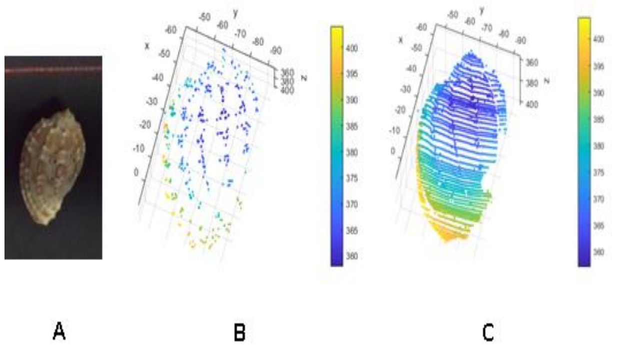


Figure 1. The results of our framework. SLAM-based laser scanning produces a semi-dense 3D model. A: laser scanning process image; B: a sparse point cloud by ORB-SLAM2; C: a semi-dense point cloud by our method.

According the method of SLAM-based laser scanning, our experiments were performed on a video sequence of real object model, where the binocular camera and the line laser projector had a fixed displacement with each other as one rig. This rig slowly swept across the target object. Firstly, the laser triangulation method is used to calculate the 3D geometric information of the laser line on each frame of image. It combines with the existing stereo SLAM method [11][12][13]to perform camera pose estimation. The laser lines on different frames are converted to the same angle of view according to the pose transformation. Then a semi-dense three-dimensional point cloud can be spliced. Secondly, in order to make the reconstruction result closer to the true scale, we introduce the mapping factor and combine the point cloud of SLAM with the point cloud result of laser scanning. Our approach effectively simplifies equipment requirements while ensuring reconstruction accuracy through efficient calculations. We experimented with objects in different shapes and textures. The experimental results show that our method can obtain semi-dense and high-precision reconstruction results with less computational and system costs.

2. RELATED WORK

This paper addresses the problem of shape recovery from multi-view images. The proposed method combines two 3D reconstruction strategies of SLAM and laser scanning.

1)SLAM: Simultaneous Localization and Mapping (SLAM) has been a hot research topic in the last two decades in Computer Vision. Place recognition is a key module of a SLAM system to close loops. Visual SLAM can perform only using a monocular camera, which is a cheapest and smallest sensor setup. However, as depth can be never fully perceived from just one camera, the scale of the depth map and estimated trajectory is arbitrary. Additionally, monocular SLAM[14][15]suffers from scale drift and may fail with pure rotations in exploration. The dense stereo SLAM method can solve all the above problems. However, their high dimensions make them computationally expensive and not suitable for strict probabilistic reasoning. Our method achieves semi-dense surface reconstruction by using the precise positioning of the sparse stereo SLAM combined with high-precision laser scanning.

2)Laser triangulation: Laser scanning sensors, which can provide high frequency and accurate range measurements. They are widely used in onboard environment sensing and navigation of mobile robots, including 3D laser scanners[16][17]. The core part is the laser triangulation method. Laser triangulation method is a three-dimensional reconstruction method solved by the principle of triangulation. Triangulation is a 3D positioning problem that determines a point based on a collection of

November-December 2019
ISSN: 0193-4120 Page No. 949 - 955

corresponding image locations and a known camera. With laser triangulation method, the laser projector can be regarded as a camera and form a triangle relationship with the other camera. Compared with the binocular triangle principle, the laser triangulation method is simpler and more precise. The recognition of the laser structured light is easier to match with respect to the image feature points. Because the depth of the laser is very reliable, researchers often use it as a baseline for error corrections. For example, the laser line in[18] was used to correct the deviation of the luminosity stereo, which could make the reconstruction details closer to the true value.

3. METHOD

3.1 Overview of method

The pipeline of our approach is shown in Figure 2. At the beginning of the workflow, the camera internal parameters are firstly calibrated according to the method in [19]. The images are captured using a binocular camera and rectified with the obtained camera parameters for distortion correction. Laser triangulation [20] is used to calibrate the laser plane equation. A semi-dense reconstruction consisting of two sequential modules follows. In the first module, stereo ORB-SLAM2[11]is used to obtain the camera pose for each frame of image and record the pose transformation matrix. Then the three-dimensional point coordinates of the laser line are extracted by laser triangulation calculation. The laser line is transformed to the same world coordinate system using the calculated pose transformation matrix. The semi-dense point cloud of the target object is spliced. In the second module, to make the reconstruction result closer to the real scale, a mapping factor is introduced to fuse the point cloud from SLAM and the laser scanning point cloud in the same scale. An optimized 3D model is finally achieved.

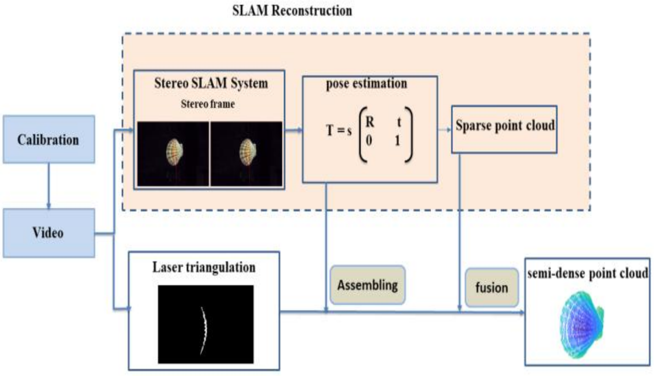


Figure 2. Overview of the proposed method.

3.2 Laser point cloud stitching

1)Review of laser triangulation: Taking the line laser triangulation method used in this paper as an example, one binocular camera and a line laser projector constitute a triangle relationship. If we define the camera coordinate system as a world coordinate system, there will be a relationship between the camera normalized coordinate system and the world coordinate system as follows:

$$\frac{x}{X} = \frac{y}{Y} = \frac{z}{Z}$$

where (X, Y, Z) is the target 3D point on the world coordinate system, and (x, y) is the image 2D point on the corresponding normalized coordinate system. Assuming that the camera has been calibrated, an association between the physical imaging plane coordinate system and the normalized image plane coordinate system can be established.

In addition, the laser plane generated is known and expressed as a plane expression as:

$$AX + BY + CZ + D = 0$$

where A, B, C and D are the coefficients of the plane expression. The laser projection point on the target object is the intersection between the plane expressed in Eq. (2) and the camera projection ray. This can be expressed as (3). Thereby, the real three-dimensional coordinate of the laser projection point can be obtained.

Published by: The Mattingley Publishing Co., Inc.

November-December 2019
ISSN: 0193-4120 Page No. 949 - 955

$$\begin{cases} X = \frac{-Dx}{Ax+By+C} \\ Y = \frac{-Dy}{Ax+By+C} \\ Z = \frac{-D}{Ax+By+C} \end{cases} \quad (3)$$

2) Calculate the transformation matrix: The laser video sequence can be acquired by the binocular camera during the reconstruction process. In each frame with estimated camera, the laser triangulation method is used to extract the coordinate information of the laser line in the frame. In the process of acquiring video sequences, stereo SLAM [11]is utilized to provide an accurate estimate of the camera pose. In particular, SLAM based on feature point is used as the baseline as described in[11]. It is widely accepted that [11] is currently the most robust method and the stereo version can provide a true scale in the result. It usually uses the first frame as the initial frame to establish the world coordinate system. The estimated camera pose is described relatively to the first frame to provide the rotation translation matrix:

$$G = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \textit{ with } R\in SO(3) \textit{ and } t\in\mathbb{R}^3 \quad (4)$$

where R is the rotation matrix, t is the translation vector, and $\mathbb{R}^3$ is the Lie group. Due to that SLAM is measured in meter, and laser scanning is described in millimeter, different from the original ORB-SLAM2, we adjust the camera's translation matrix by introducing a scale factor s relative to the original camera pose to restore the transformation at a true scale. The pose refinement based on the pose map optimization is defined as:

$$T = s\begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \textit{ with } R\in SO(3),\ t\in\mathbb{R}^3 \textit{ and } s\in\mathbb{R}^+ \quad (5)$$

3) Point cloud integration: We use the pose transformation equation to splice the laser lines into the same world coordinate system. The equation is as follows:

$$P_W = T \cdot P_C \quad (6)$$

where P_W is the position information in a world coordinate system, and P_c is the position information in a camera coordinate system of each frame. The above operation is performed for each frame, so that the three-dimensional semi-dense point cloud of the target object can be spliced.

3.3 Equal-scale mapping

Dueto the world coordinate system initialized by SLAM differs from the world coordinate system defined by the laser triangulation method for calculating the position of the laser point, the target object spliced by the laser point cloud has a deviation in shape. Therefore, the mapping factor is introduced in this paper. The SLAM point cloud and the laser scanning point cloud are scaled together for deviation correction, and finally a semi-dense three-dimensional model with high precision and real scale is generated. The following is a specific automatic iterative process.

Step 1: Adjust the ORB-SLAM2[11] and the internal parameters of the laser reconstruction to make the distortion consistent. All the pictures are processed based on the correction picture.

Step 2: Set the mapping factor β. We use the point cloud of SLAM and laser scanning to initialize the mapping factor, as follows:

$$\beta_0 = \frac{\frac{1}{n_s}\sum_{i=1}^{n_s} d_s^i}{\frac{1}{n_l}\sum_{j=1}^{n_l} d_l^j}$$

where $n_s$and n_l represent the number of points of the slam point cloud and the laser scanning point cloud respectively, d_s and d_l represent the average depth of the slam point cloud and the laser scanning point cloud respectively.

Step 3: Return the mapping factor to the laser line splicing part and re-splicing. It restores a new point cloud, and records the mapping factor, as show in (8):

Published by: The Mattingley Publishing Co., Inc.

$$\beta_w = \frac{\frac{1}{n_s}\sum_{i=1}^{n_s} d_s^i}{\frac{1}{n_l}\sum_{j=1}^{n_l} d_l^j}\cdot\beta_{w-1} \tag{8}$$

where β_w represent the result of w iterations.

Step 4: If β converges to 1, the whole process is ended. Otherwise, jump to step 3 to continue.

Through the adjustment of the above iterations, the ORB-SLAM2[11] point cloud can be merged with the laser scan results. The result of the fusion is back-projected onto raw image without significant deviation. As shown in the figure 3 below, the left side is the result of the back-projection of the point cloud without adjustment, and the right side is the result of the back-projection of the point cloud after the introduction of the mapping factor. By comparison, our method significantly reduces the bias.

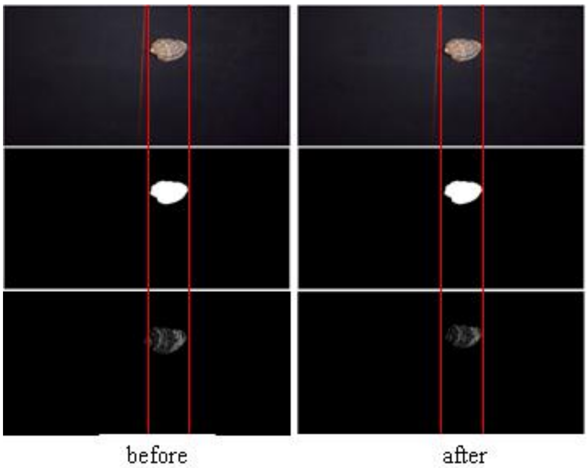


Figure 3. Comparison of back-projection results.

4. EXPERIMENTS

4.1 Experimental equipment setup

We evaluate our approach using real data sets. The imaging setup is used for 3D reconstruction of the actual object(Figure 4). The high-definition binocular camera and the line laser are located on the same plane, and it can be smoothly placed on the hand-held support frame. The support rod can also be used as a device handle. The camera is at the center of the handle and the laser line slowly sweeps over the target while the image is being captured. Note that before the video capture, the calibration of the binocular camera and the calibration of the laser plane equation are performed firstly. In the following sections, we will present a detailed reconstruction of various real world objects.

Figure 4. Imaging setups for real world object capture.

4.2 Qualitative Analysis

Our method was tested on several real objects, including shells, monkeys, conch and small balls. We used a self-made experimental setup to collect scanned images of each object for surface repair.

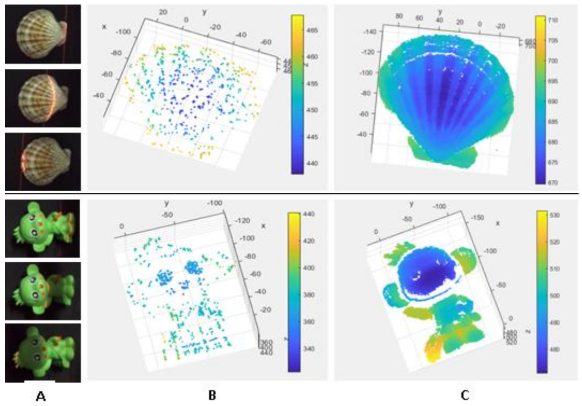
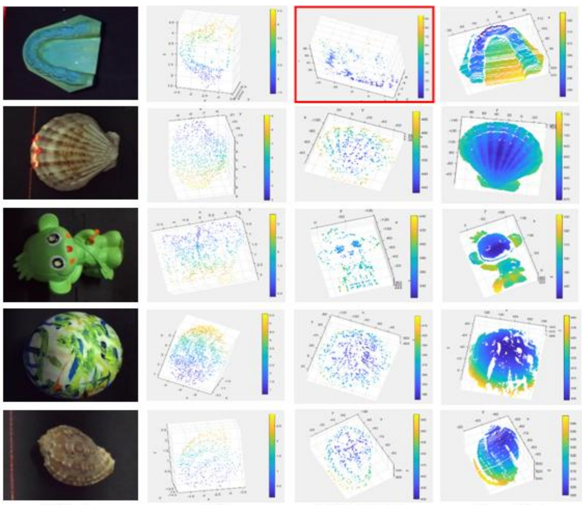


Figure 5. Reconstruction results for shells and monkeys. (A) Input image (part of the image scanned by the laser line). (B) Sparse points calculated by ORB-SLAM1. (C) 3D shape recovered by our method.

Figure 5 shows the reconstruction results of the shell and monkey. These results were captured by settings based on binocular cameras. In the experiment, the object was captured about 450 mm from the camera. The image resolution is 1920×1080. A of Fig. 5 shows a representative input image. Then we reconstruct the sparse point cloud using ORB-SLAM2 and calculate the pose (in B of Figure 5). Based on the calculated camera pose, we first spliced the results of the laser scan, and then used the introduced mapping factor to fuse the ORB-SLAM2 point cloud and laser scan results. C of Figure 4 shows the shape reconstructed by our method. We effectively combine the two and get the detailed shape through the details. The texture on the shell and monkey is clearly visible.

In addition, in order to prove the effectiveness of the SLAM-based laser scanning method, we compared it with some traditional methods. As mentioned earlier, we compared it with the SFM and SLAM methods. The results are shown in Figure 6. Figure 6 shows the comparison of tooth models, shells, conch, toy monkey and small ball, which were taken by our camera. The first column is one of the input images as a reference frame for shape reconstruction. The second column shows the reconstructed point cloud through the SFM[3]. For a limited input image (approximately 15 images), the SFM-based point cloud produces very sparse results. The third column of Figure 6 shows the reconstruction results using ORB-SLAM2. It can be seen that the point cloud is still sparse, but the shape of the object is roughly the same as the real object. The last column is reconstructed by our method, our method has higher precision, the result is denser, and there is a clearer texture.



Objects SFM ORB-SLAM2 Our method

Figure 5. The left is the objects we shot. The three columns on the right are reconstruction results of SFM, ORB-SLAM2, and our method respectively. Note that in the first rows, the objects marked with a red rectangle is a failed reconstruction. The reason is that the information collected by ORB-SLAM2 is too sparse to get the reconstruction result.

5. CONCLUSION

We have shown how we combine Simultaneous Localization and Mapping (SLAM) with laser scanning for semi-dense 3D reconstruction. Experimental results demonstrate the effectiveness of our method. In practical applications such as 3D printing and augmented reality (AR), a more elaborate 3D model with fine-grained information is needed. In the future work, we will consider recovering a dense high-resolution 3D model.

REFERENCES

[1] C. Hong, J. Yu, D. Tao, and M. Wang, “Image-based three-dimensional human pose recovery by multiview locality-sensitive sparse retrieval,” IEEE Transactions on Industrial Electronics, vol. 62, no. 6, pp. 3742– 3751, 2015.

[2] L. Zhang, M. Wang, W. Li, R. Hong, and M. Liu, “An automatic threedimensionalscenereconstructionsystem usingcrowdsourcedgeo-tagged videos,” IEEE Transactions on Industrial Electronics, vol. 62, no. 9, pp. 5738–5746, 2015.

[3] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski, “Building rome in a day,” in 2009 IEEE 12th international conference on computer vision. IEEE, 2009, pp. 72–79.

[4] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski, “Towards internetscale multi-view stereo,” in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pp. 1434–1441. IEEE, 2010.

[5] C. Roman, G. Inglis, and J. Rutter, “Application of structured light imaging for high resolution mappingof underwater archaeological sites,” in OCEANS’10 IEEE

November-December 2019
ISSN: 0193-4120 Page No. 949 - 955

SYDNEY. IEEE, 2010, pp. 1–9.W.-K. Chen, Linear Networks and Systems (Book style).Belmont, CA: Wadsworth, 1993, pp. 123–135.

[6] A. J. Davison, “Real-time simultaneous localisation and mapping with a single camera.” in Iccv, vol. 3, 2003, pp. 1403–1410.

[7] G. Klein and D. Murray, “Parallel tracking and mapping for smallar workspaces,” in Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. IEEE Computer Society, 2007, pp. 1–10.

[8] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, “Dtam: Dense tracking and mapping in real-time,” in 2011 international conference on computer vision. IEEE, 2011, pp. 2320–2327.

[9] T. Whelan, R. F. Salas-Moreno, B. Glocker, A. J. Davison, and S. Leutenegger, “Elasticfusion: Real-time dense slam and light source estimation,” The International Journal of Robotics Research, vol. 35, no. 14, pp. 1697–1716, 2016.

[10] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, “Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments,” The International Journal of Robotics Research, vol. 31, no. 5, pp. 647–663, 2012.

[11] R. Mur-Artal and J. D. Tard´os, “Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras,” IEEE Transactions on Robotics, vol. 33, no. 5, pp. 1255–1262, 2017.

[12] J. Engel, J. St¨uckler, and D. Cremers, “Large-scale direct slam with stereo cameras,” in 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2015, pp. 1935–1942.

[13] R. Wang, M. Schworer, and D. Cremers, “Stereo dso: Large-scale direct sparse visual odometry with stereo cameras,” in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 3903–3911.

[14] A.Davison, I.Reid, N.Molton, and O.Stasse. Mono SLAM: Real-time single camera SLAM. Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 29, 2007. 2

[15] J. Engel, T. Sch¨ops, and D. Cremers. LSD-SLAM: Largescale direct monocular SLAM. In European Conference on Computer Vision (ECCV), 2014. 2, 11

[16] P. K. Allen, I. Stamos, A. Gueorguiev, E. Gold, and P. Blaer, “Avenue: Automated site modeling in urban environments,” in International Conference on 3-d Digital Imaging and Modeling, 2001.

[17] M. Hebert, M. Deans, D. Huber, B. Nabbe, and N. Vandapel, “Progress in3–dmappingandlocalization,”in Proceedings of the 9th International Symposium on Intelligent Robotic Systems,(SIRS01). Citeseer, 2001.

[18] F. Hao, Q. Lin, W. Nan, J. Dong, and Y. Hui, “Deviation correction method for close-range photometric stereo with nonuniform illumination,” Optical Engineering, vol. 56, no. 10, p. 1, 2017.

[19] Zhang, Zhengyou. "A flexible new technique for camera calibration." IEEE Transactions on pattern analysis and machine intelligence 22 (2000).

[20] H. Zhu, L. Lu, L. Qi, and J. Dong, “An automatic laser sheet calibration method for laser scan range measurement system,” in Tenth International Conference on Graphics and Image Processing (ICGIP 2018), vol. 11069. International Society for Optics and Photonics, 2019, p. 1106939