Proceedings of the 2016 IEEE
International Conference on Robotics and Biomimetics
Qingdao, China, December 3-7, 2016

# Visual and LiDAR-based for The Mobile 3D Mapping

Qiao Wu[1], Kai Sun[2], Wenjun Zhang[2], Chaobing Huang [1*]， Xiaochun Wu[3]

*Abstract*—**Multi-sensors fusion to construct an accurate 3D map about unknown indoor scene and outdoor in the simultaneous localization and mapping (SLAM) area is becoming increasingly popular. In this paper our methods pay an attention to how to optimize the 2D map from the indoor or outdoor unknown large-scale environment and how to make the 3D map intuitive and accurate. Firstly, this paper accomplished the calibration between 2D Light Detection And Ranging sensor （LiDAR） and panoramic camera, using horizontal 2D LiDAR to finish 2D SLAM. Then the new method that closed-loop detection was based on visual was used to assist the 2D SLAM to get a better optimization result in sub-maps efficiently. Finally, vertical LiDAR data was added to get the 3D point cloud, using the images from a panoramic camera to accomplish the registration with point cloud. The result of final 3D map with RGB information demonstrated that our methods are feasible.**

## I. INTRODUCTION

3D mapping in an indoor scene can help us view the three-dimensional of objects and spatial structure on our computer efficiently. We can even have an accurate measurement about the objects . All of these, we need to have a precise description about the scene through the sensors. In recent years, simultaneous localization and mapping (SLAM) has become more and more popular ,there are three major categories :visual-based and laser-based and RGBD-based. The visual-based SLAM like ORB-SLAM [1]and LSD-SLAM [2],they only use the image information to construct the 3D space , they need overcome the scale-drift problem[3] and they can not overcome the scene like few features or the lighting change quickly. the RGBD –SLAM [4] [5] [6] use the sensor like Kinect, Xtion Pro. Their scan angle and distance are limited to be small, the workload will be heavy if we want to get the point cloud of a large building. Owing to the different sensors the laser-based SLAM can be divided into 2D SLAM [7] [8] [9] [10] and 3D SLAM [11] .Certainly the 2D LiDAR also can be used for constructing the 3D environment [12] [13]. The laser-based SLAM can work well at different environments. Since the laser can reach far away, we can use the laser-based SLAM at large-scale scene and get dense enough point cloud data.

The 2D laser-based SLAM have three paradigms, which are Kalman filters, Particle filters and Graph-based. Kalman filters and Particle filters are two implementations of Bayes filters[10]. The Kalman filters have two main parts: prediction and update. In order to solve the nonlinear problem, the Extended Kalman Filter (EKF) was put forward. The EKF-SLAM can only deal with a single mode, it is successful in medium-scale scene but when it comes to a large map, it becomes computationally intractable. Certainly, there is better method solving the nonlinear problem like Unscented Kalman Filter (UKF). The Kalman filter and its variants can only model Gaussian distributions, so we need an approach to deal with the arbitrary distributions, Particle filters can deal with the arbitrary distributions by using multiple samples. This method deems that the more particles fall into a region, the higher the probability of the region [14]. The posterior probability is represented by a set of particles which has been weighted. The particle filters-based SLAM like [7], model the robot's path by sampling and computing the landmarks given the path. Graph-based SLAM [9] considers that a graph is composed of poses and constraints between poses. By constructing a graph to minimizes the sum of the squared error, in fact, it is a method of optimization that use linear methods to solve the non-linear problem .

In this paper we use the Rao-Blackwellized particle filters [15] to finish the 2D SLAM without the GPS or IMU in indoor or relatively flat outdoor scene, this method based on recursive Bayes filter. In the Factorization of the SLAM posterior:

$$p(x_{1:t}, m|z_{1:t}) = p(m|x_{1:t}, z_{1:t})p(x_{1:t}|z_{1:t}) \quad (1)$$

In this formula, $x_{1:t}$ represents the path of robot, $z_{1:t}$ represents the observations, $m$ represents the map. In order to obtain the map, the potential trajectory of robot and the observations are required, the process of Rao-Blackwellized particle filters can be divided into four parts :

Sample: based on the proposed distribution $\pi$ , get the next generation particles from the current particles.

Calculate importance weights:

$$w_t^{[i]} = \frac{p(x_t^{[i]}|x_{1:t-1}^{[i]}, z_{1:t})}{\pi(x_t^{[i]}|x_{1:t-1}^{[i]}, z_{1:t})} \quad (2)$$

Resample: replace the low importance weights particles with the high.

*This work is supported by the National Science Foundation of China (NSFC) through Grant No. 51579204

[1] Qiao Wu is with the Key Laboratory of Fiber Optic Sensing Technology and Information Processing, Ministry of Education,Wuhan University of Technology, Wuhan, Hubei, 430070, China (e-mail: wuqiao@ whut.edu.cn).

[2]Kai Sun is with the Leador Spatial Information Technology Corporation, Wuhan, 430000 ,China (e-mail: sunkai@leador.com.cn).

[2]Wenjun Zhang is with the Leador Spatial Information Technology Corporation, Wuhan, 430000 China (e-mail:zhangwenjun@leador.com.cn).

[1*] Chaobing Huang is with the Key Laboratory of Fiber Optic Sensing Technology and Information Processing, Ministry of Education,Wuhan University of Technology, Wuhan, Hubei, 430070, China (e-mail:huangcb@ whut.edu.cn).

[3]Xiaochun Wu is with the school of Navigation, Wuhan University of Technology, Wuhan, hubei ,China . ( e-mail:xcwu@whut.edu.cn).

Map evaluate: update each particle map by using the known pose mapping.

In this paper, we use two 2D LiDAR with scan angle of $270^{\circ}$ to get the point cloud of 3D environment, they are perpendicular to each other. The panoramic camera is used to get the image. Our experimental platform is shown in figure 1.

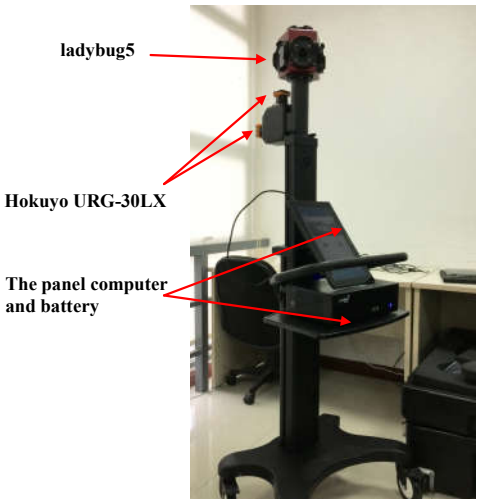



Figure 1. The mobile experimental platform provided by Leador Spatial Information Technology Corporation, it mainly consist of two 2D LiDAR (Hokuyo URG-30LX) and a high resolution panoramic camera(ladybug5), these equipments powered by lithium batteries (12V, 36Ah) , the panel computer can be operated to obtain the data. The fastest speed of the platform when it works is 3m/s.

## II. Procedure For 3D Mapping

The procedures to get indoor environment's 3D map are composed of the calibration between two LiDARs with the panoramic camera, getting raw images and horizontal LiDAR data ,using the images from the panoramic camera's frontal camera to detect the loop-closure ,using the horizontal LiDAR data to accomplish the LiDAR odometry and 2D sub-maps, Various optimization methods are implemented to ensure the accuracy of the map, then vertical LiDAR data is added to get the 3D map .In the end the registration between the panoramic image and the point cloud to get the RGB 3D map.

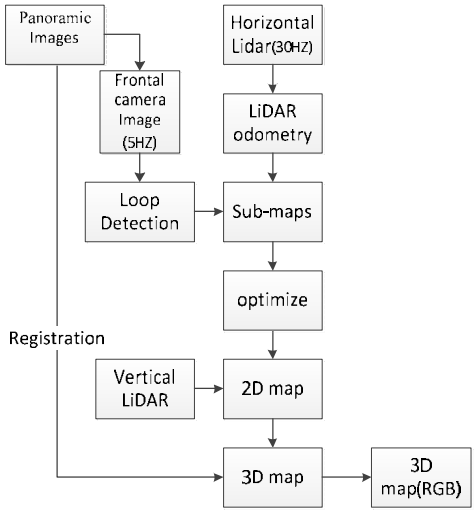


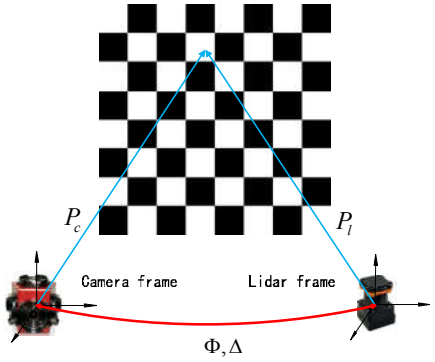

Figure 2. Block diagram of the 3D mapping

## III. Our Method

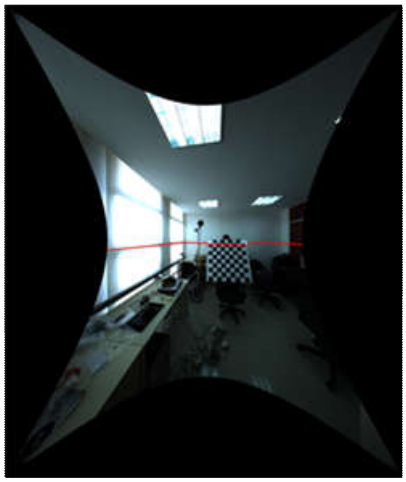
### A. Calibration

The 2D LiDAR can obtain the range information but can not get the texture and color information , which is very important for object detection and recognition[16][17], the camera can do these work. in order to get the complete information to have an accurate 3D model of the environment. We associate their data by calibration.

The calibration can get the relative transform between the camera and 2D the LiDAR. Although we can not see the laser line, the back-projection from LiDAR point cloud can be seen on the panoramic image after we know the extrinsic parameters.




(a)



(b)

Figure 3. In the (a), the calibration to determine the extrinsic parameters between the LiDAR and the camera, the $\Phi$ and $\Delta$ denote the coordinate transformation from the LiDAR frame to the camera frame, the $P_c$ and $P_l$ are the points define on the checkerboard from the camera and the LiDAR frame. In the (b), the red line is the back-projection result which show the laser point emitted from 2D LiDAR ,the image is from the frontal camera on the ladybug5.

### B. Create 2D Sub-maps

Instead of creating a large map, in this paper, map is managed by sub-maps like the [18] but different. Our method steps are as follows:

1) We define the global map space size is N*N, the center of the space is (N/2, N/2), a global map space is consisted of many sub-maps, we define a sub-map space is n*n;
2) When we have a map point, based on its coordinate relative to the center point, we apply a space (n*n) and set an initial value to grid cell in the corresponding sub-map. The value in this grid cell will be increased when there are later map points falling into this grid cell.
3) In the end, we get each sub-map' starting and ending coordinate.

Because the method of particle filter need estimate the states and the observations constantly. Compared with a large map, this method can reduce the accumulated error caused by inaccurate observations and states after long-term move in large-scale scene, each sub-map uses their features independently and each sub-map we give their reasonable space, the new space will be given after the judgment whether or not new point outside its forward spaces, so this method save the memory space more effectively than allocate the sufficient space to manage the map about unknown environment directly. Meanwhile, when we load the global map on our software, we can make multiple sub-maps with simultaneous loading.

The method of grid occupied is used to build sub-maps, we just concerned about the outline of buildings or obstacles. So when single-line laser meet the obstacles on one point, this point is occupied. We define one grid point $m$ , o denotes this point is occupied , $t$ denotes the times this point is scanned. $O_C$ denotes the possibility of this point was occupied and the $O_n$ is the possibility this point not be occupied. $c_n$ and $c_o$ denote the number of occupied and free times respectively.

$$O\left(o\left|x_{1:t},z_{1:t}\right.\right)=\prod_{i=1}^{t}O\left(o\left|x_t,z_t\right.\right) \tag{3}$$

Logarithmic form can be written as:

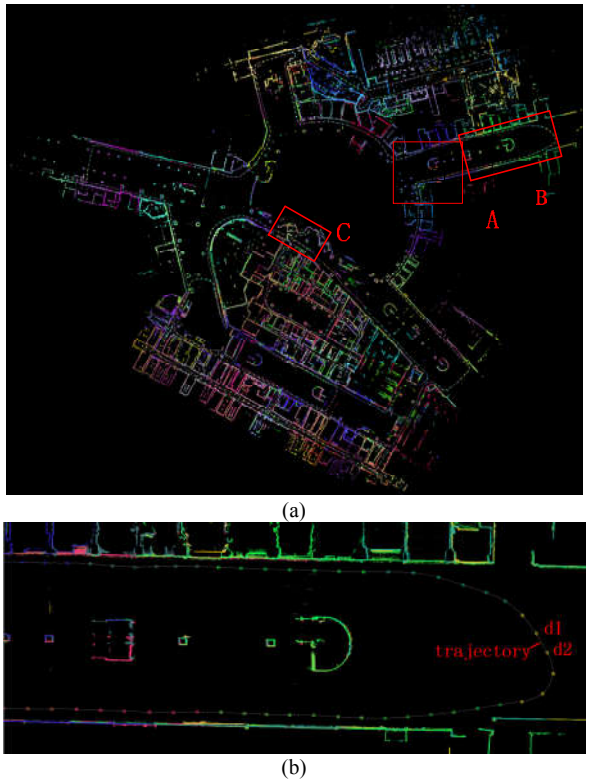
$$\log O\left(o\left|x_{1:t},z_{1:t}\right.\right)=\log O_n^{c_n}+\log O_o^{c_o} \tag{4}$$

In the end, the occupied possibility of $m$ can be written as:

$$p(o\left|x_{1:t},z_{1:t}\right.)=\frac{\exp(c_n\log O_n+c_o\log O_o)}{1+\exp(c_n\log O_n+c_o\log O_o)} \tag{5}$$

So the $p(m\left|x_{1:t},z_{1:t}\right.)$ is equal with $p(o\left|x_{1:t},z_{1:t}\right.)$ , the occupied point ,namely the one point on the map. When we get each sub-map's $p(x_{1:t}\left|z_{1:t}\right.)$ , we get the result of 2D SLAM. The framework of sub-maps makes the offline loading and handling the map more convenient and efficient.

Using our platform, we completed the data collection at Guanggu Walking Street, Wuhan, Hubei Province, China. This area covered about 20000 square meters, indoor and outdoor environment are included. This 2D map was composed of 173 sub-maps which we use different colors to express after binaryzation based on the given threshold. The trajectory, the position of the panoramic images, the outline of the buildings and obstacles are included. Meanwhile, the panoramic images and the vertical LiDAR data were also collected.




(a)

(b)

Figure 4. (a) is the result of 2D SLAM, (b) comes from the area of B in (a), d1 and d2 represent the different position of panoramic camera when the platform moves, the trajectory is the result of horizontal LiDAR odometry.

### C. Optimization

In our method, the accuracy of 2D map is closely related to the accuracy of 3D map, so it is very important to optimize the result of 2D SLAM. Especially in large-scale environment of mapping, due to the inaccurate states and observations, the accumulated errors can not be avoided completely through the method of sub-maps. These errors can make the robot's position estimate diverge. Thus global optimization based on graph is used to solve the nonlinear least squares problems [17]. The graph can be described as:

$$G=\{V,E\} \tag{6}$$

In our SLAM, $V$ represents the vertex $x_i$ or landmarks, $E$ represents the constraint $z_{ij}$ between $x_i$ and $x_j$ , $\Omega_{ij}$ represents the information matrix which related to $x_i$ and $x_j$ . $e(x_i,x_j,z_{ij})$ describes the degree of satisfaction that the $x_i,x_j$ with the constraint $z_{ij}$ , we need get the minimum of a function of this form :

$$F(X)=\sum_{<i,j>\in C}e(x_i,x_j,z_{ij})^T\Omega_{ij}e(x_i,x_j,z_{ij}) \tag{7}$$

$$X^* = \arg\min_X F(X) \tag{8}$$

We consider the problem the optimization problems as adding constrains to realize the nonlinear least squares. The location problem is deemed to focus on identification whether or not a place has been visited, but the SLAM problem, except that, the recognition of the area which has been mapped can not be ignored, it helps refine the location. Loop-closure detection in SLAM is regarded as a good solution to recognize the place which has been visited previously. it can be considered as a constraint to get a loop and make the trajectory consistent . It reduces the uncertainty in robot and landmark estimate. The two methods of detecting the loop-closure are vision based and range based. For example, [1] utilizes the image streams to extract the key frames and set the similarity threshold to determine whether two frames are in the same place. [20] uses a machine learning algorithm for training the range data into a classifier to detect loop-closure.

In this paper, In order to create 3D RGB map, our panoramic camera and 2D LiDAR data record timestamps respectively, their timestamps have an accurate correspondence. Therefore, we use the new method to rely on the frontal forward camera on panoramic camera to detect the loop-closure and seek the correspondence between their timestamps.

Based on the key frames' similarity, the ORB-SLAM had proved that the problem of detecting the loop-closure could be solved successfully by visual information, Mur-Artal *et al.* used the ORB descriptors to create the visual vocabulary they grouped keyframes in the covisibility graph .in the end they got the keyframes whose scores were higher than others, then confirmed the loop candidate from the consecutive three candidates in the covisibility graph [1] . Refer to their methods, we extracted the frontal camera image at the frequency of 5HZ, get the LiDAR data at the frequency of 30HZ, Meanwhile, we deemed each image as a keyframe due to the low frequency of the frontal camera. If we could get the time of loop-closure in the images, we could extract the LiDAR data set at the same time to decide a loop candidate.

AS is shown in the figure 5, we assume that the similarity of m1 and m2 can meet the condition of loop-closure by detecting the similarity of them. Then we consider they are at same place. In order to increase the reliability of the LiDAR loop-closure, we find the LiDAR location data set which contains location nodes of LiDAR.

$$L^s_{t2} = \{...l_{t2-2\Delta t}, l_{t2-\Delta t}, l_{t2}, l_{t2+\Delta t}, l_{t2+2\Delta t}...\} \tag{9}$$

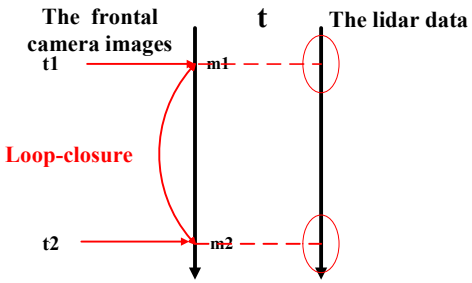
$$L^s_{t1} = \{...l_{t1-2\Delta t}, l_{t1-\Delta t}, l_{t1}, l_{t1+\Delta t}, l_{t1+2\Delta t}...\} \tag{10}$$

Then we get the two nodes which has the minimum distance between the $l_{ts} \in L^s_{t1}$ and $l_{te} \in L^s_{t2}$, $T_{threshold}$ need be decided based on the scene, consider the similarity of adjacent images ,the time must met :

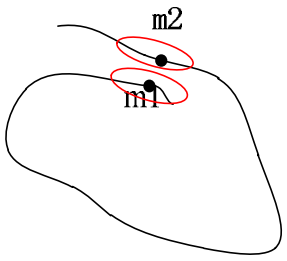
$$\Delta T_{\min} = (t2 - t1) > T_{threshold} \tag{11}$$

$$L_{d\min} = \{l_{ts}, l_{te}\} = \{l_t \in nodes \,|\, dis_{\min}(L^s_{t2}, L^s_{t1}),\ \Delta T_{\min}\} \tag{12}$$

In our experiment , we define s is 11 .Because when we get one frame of the frontal camera ,we get the 6 frames of the LiDAR location data, $L^s_{t2} = \{l_{t2-5\Delta t}, ...l_{t2}..., l_{t2+5\Delta t}\}$ .After we get the two point set $L_{d\min} = \{l_{ts}, l_{te}\}$ , we can get the pose of two points on the LiDAR odometry. We use the Iterative Closest Point (ICP) algorithm [21] to calculate the rotation matrix R and the translation T matrix, we consider the LiDAR data set between the time *ts* and *te* as a loop-closure data. About visual-based loop-closure we use the bag of word [22] (BOW2), this method is suitable for the environment, which has good lighting and the features can be extracted easily.



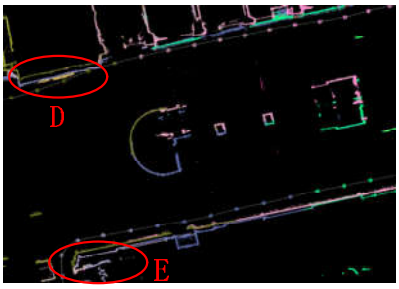

(a)



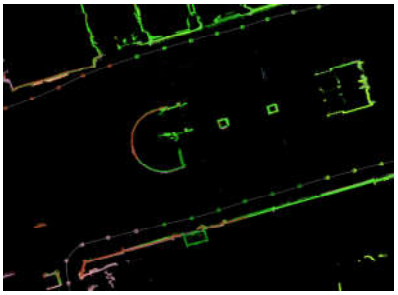

(b)

Figure 5. Visual and LiDAR-based to detect the loop-closure

After 2D SLAM, we can see the same place have different maps at different times before optimization , after detection of loop-closure and optimization , the result is improved evidently .




(a)



(b)

Figure 6. The area of D and E comes from the area of A in the figure 4(a). Before optimization is (a), after optimization is (b).

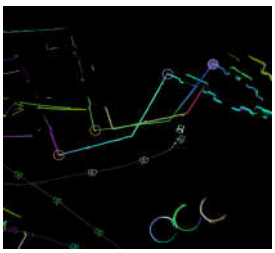
In different environments , for example ,where there is no loop-closure at local map, or our visual information can not get enough features to accomplish the detection of loop-closure, virtual landmarks are used to add the constrains to make the 2D map become accurate at the offline process instead of placing the markers in real scenes. We add the vertexes and the edges in a graph by adding the virtual landmarks in our sub-maps. This method can optimize the LiDAR odometry gradually. In fact it can be understood to realize the maximum posterior probability estimation $X^*$, which is the location of experimental platform and the landmarks. $X$ represents the state positions , $L$ represents the landmarks , $Z$ represents the constraints between the landmarks and state positions or between the states positions.

$$X^* = \arg\max_X p(X, L|Z) \tag{13}$$

$$p(X, L|Z) \propto \prod P(x_{i+1}|x_i, z_{ii+1}) \cdot \prod p\left(l_j|x_i, z_{ij}\right) \tag{14}$$

$\prod p(x_{i+1}|x_i, z_{ii+1})$ represents the odometry constraints,

$\prod p\left(l_j|x_i, z_{ij}\right)$ represents the landmarks constraints.



(a)

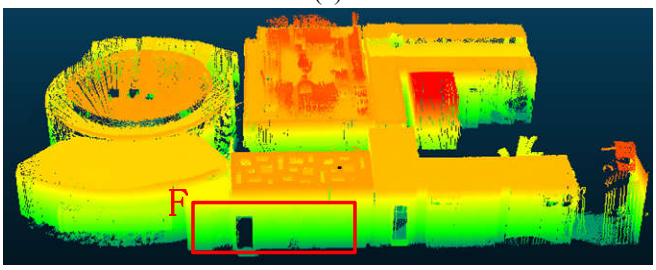
(b) (c)

Figure 7. (a),(b),(c) all come from the area C in the figure 4 (a).By adding the landmarks gradually from (a) to (c), the real outline of buildings become more clearly.

## IV. Experiment and Results

In the process of experiment, we chose a relative small scene to show our final integral results, this area covered about 300 square meters. Fusing the vertical and horizontal point cloud .After calibration, we could get the correspondence between the panoramic camera image pixels with the point cloud data, based on the extrinsic parameters, we fused the image with the point cloud successfully, then we could get the 3D RGB map.

(a)




(b)

(c)

(d)

(e)

Figure 8. (a) is the result of after optimization which contains the LiDAR odometry and the 2D map ,(b) is the height ramp of the original point cloud, (c)is the area of F internal structure in the (b) ,the (e) is the colored point cloud after data fusion ,the (d) is one part of the (e) at same place with (c).

The accuracy of the 3D data is closely related to the data from the optimization of 2D SLAM, In order to verify the accuracy, we compared the horizontal distance value from the Electronic Total Station with the distance data we measured on our software, there were four targets In the figure 8(a), we defined those four targets were *t1, t2, t3, t4* from top to bottom. We thought the horizontal distance data from the Electronic Total Station was the real distance in the real scene. In order to get a more accurate measurement on our software, we took the mean value of ten measurements as the final result.

Table1 illustrates the collected data got by our experimental platform could met the accuracy requirements within 5cm.

TABLE I. THE HORIZONTAL DISTANCE

| **Target–Another Target** | **The Distance data from Electronic Total Station(m)** | **The Distance data from Our Collected Data On Our Software(m)** |
|---|---|---|
| *t1-t2* | 12.073 | 12.062 |
| *t1-t3* | 18.386 | 18.364 |
| *t1-t3* | 25.998 | 25.970 |
| *t2-t3* | 6.445 | 6.441 |
| *t2-t4* | 14.961 | 14.942 |

## V. CONCLUSION AND FUTURE WORK

In this paper, by calibration we got the correspondence between the image and point cloud. Sub-maps and optimization methods were used to make the accumulate error as small as possible. We used a simple but effective new method to find the LiDAR loop-closure and got better optimization results. In the end we got the 3D RGB map.

Our experimental platform is mainly to deal with indoor scene mapping, to solve the problem of location in indoor scene which has no GPS signal or relative flat outdoor scene. It does not rely on the GPS and IMU to locate quickly and get the colorful point cloud by offline data fusion and optimization. The platform can be used to build indoor 3D colorful point cloud in a variety of modern facilities in public buildings, factories, shops, schools and so on. The scan data can be processed by software to get the 3D colorful point cloud data, 5cm accuracy can be ensured in less than one thousand square meters of indoor scene and 10cm accuracy can be ensured in large-scale scene. The complexity of the data processing is related to the complexity and area of the scene. Our point clouds data can be measured precisely. In the future, we want to make our experimental platform combine with IMU and GPS. By performing this method, our experimental platform will be applicable to more complex scene.

## ACKNOWLEDGMENT

The author would like to thank three colleagues from the Leador Spatial Information Technology Corporation, Qiao Zhang, Zhen Shu, Qiang Gao for their help on source code and valuable suggestion ,discussion .

## REFERENCES

[1] Mur-Artal R, Montiel J M M, “ORB-SLAM: A Versatile and Accurate Monocular SLAM System.,” IEEE Transactions on Robotics, 2015, pp .1147-1163.

[2] Jakob Engel,Thomas Sch ¨ ops and Daniel Cremers, “LSD-SLAM: Large-scale direct monocular SLAM,” In: Computer Vision – ECCV 2014. Springer, 2014, pp. 834 – 849.

[3] Strasdat, Hauke, J. M. M. Montiel, and Andrew J, “Davison,Scale drift-aware large scale monocular SLAM,” Robotics: Science and Systems VI , 2010.

[4] Kerl C, Sturm J, Cremers D, “Dense visual SLAM for RGB-D cameras,”2013, pp.2100-2106.

[5] Fioraio N, Konolige K. Realtime Visual and Point Cloud SLAM. Proc of the Rgb, 2011.

[6] Izadi S, Kim D, Hilliges O, “Kinectfusion: real-time 3D reconstruction and interaction using a moving depth camera,”Proc Uist, 2011, pp.559--568.

[7] Montemerlo M, “FastSLAM: a factored solution to the simultaneous localization and mapping problem,” Eighteenth national conference on Artificial intelligence. American Association for Artificial Intelligence, 2002:2004.

[8] Thrun S, Montemerlo M, “The Graph SLAM Algorithm with Applications to Large-Scale Mapping of Urban Structures,” International Journal of Robotics Research, 2006, pp.403-429.

[9] Durrant-Whyte H, Roy N, Abbeel P, “A Linear Approximation for Graph-Based Simultaneous Localization and Mapping,” Robotics: Science & Systems Vii, University of Southern California, Los Angeles, Ca, Usa, June. MIT Press, 2011, pp.41 - 48.

[10] Steux B, El Hamzaoui O, “ tinySLAM: A SLAM algorithm in less than 200 lines C-language program,” International Conference on Control, Automation, Robotics and Vision, Icarcv 2010, pp.1975-1979.

[11] Pomerleau F, Krusi P, “Long-term 3D map maintenance in dynamic environments,”IEEE International Conference on Robotics & Automation. 2014, pp.3712 - 3719.

[12] Zhang J, Singh S, “LOAM: Lidar Odometry and Mapping in Real-time,”2014.

[13] Bosse, Michael, and R. Zlot, “Continuous 3D scan-matching with a spinning 2D laser,” IEEE International Conference on Robotics and Automation IEEE Press, 2009, pp.4312-4319.

[14] Thrun, Sebastian, Wolfram Burgard, and Dieter Fox. Probabilistic robotics. MIT press, 2005.

[15] Grisetti, G., C. Stachniss, and W. Burgard. “Improved Techniques for Grid Mapping With Rao-Blackwellized Particle Filters,” IEEE Transactions on Robotics.2007, pp.34-46.

[16] Kwak, Kiho, “Extrinsic calibration of a single line scanning lidar and a camera,” Ieee/rsj International Conference on Intelligent Robots and Systems IEEE, 2011, pp.3283-3289.

[17] Carlevaris-Bianco, Nicholas, “Visual localization in fused image and laser range data,” 2011, pp.4378-4385.

[18] Eliazar, Austin, and R. Parr, “DP-SLAM: Fast, Robust Simultaneous Localization and Mapping Without,”2003, pp.1135--1142.

[19] Kummerle, R., et al. "G 2 o: A general framework for graph optimization." IEEE International Conference on Robotics and Automation IEEE, 2011,pp.3607-3613.

[20] Granstrom, K.,“Learning to detect loop closure from range data,”2009, pp .15-22.

[21] Rusinkiewicz S, Levoy M, “Efficient Variants of the ICP Algorithm,3dim. IEEE Computer Society, 2001, pp.145.

[22] Galvez-López D, Tardos J D, “Bags of Binary Words for Fast Place Recognition in Image Sequences,” IEEE Transactions on Robotics, 2012, pp.1188-1197.