ORB-SLAM based ICP and Optical Flow Combination Registration Algorithm in 3D Dense Reconstruction

Liya Duan[1], Renxia Wu[2] , Xin Wang[1], Yinghe Liu[3], Yongjie Ma[1] and Hui Fang[2]
[1]*Institute of Oceanographic Instrumentation, Qilu Univercity of Technology (Shandong Academy of Sciences), No.37 Miaoling Road, Qingdao, China*
[2] *Institute of Coastal and Offshore Engineering, School of Engineering, Ocean University of China, No. 238 Songling Road, Qingdao, China*
[3] *Institute of Oceanology, Chinese Academy of Sciences, No.7 Nanhai Road, Qingdao, China*

Keywords: ICP, Optical Flow, ORB-SLAM, 3D Reconstruction .

Abstract: Dense 3D reconstruction is important for applications in many fields. The existing depth information based methods are typically constrained in their effective camera-object distance that should be from 0.4m to 4m. We present a combination of optical flow method and ICP method, and fuse this method and RGB-D camera ORB-SLAM to make full use of the color and depth data and refine the registration results that can achieve a more accurate dense 3D reconstruction. The experiment shows that this combination method is effective for the dense map acquisition.

1 INTRODUCTION

Registration Algorithm is of importance in dense 3D reconstruction with many applications in object recognition (Hong, 2015), object retrieval (Lu, 2014), scene under-standing (Ala, 2014), object tracking (Kim, 2014), autonomous navigation (Valiente, 2014), human–computer interaction (Sun, 2015), telepresence, tele- surgery, reverse engineering, virtual maintenance and visualization (Galvez, 2012). Simultaneous Localization and Mapping (SLAM) has been a hot research topic in the last two decades, and has recently attracted the attention of the researchers in computer vision related area. The ORB-SLAM is built on the main ideas of PTAM, the loop detection work of G´alvez-L´opez and Tard´os (Hong, 2015), the scale-aware loop closing method of Strasdat et. al (Strasdat, 2010) and their idea of a local covisible map for large scale operation (Strasdat, 2011), which is a new Monocular SLAM system that overcomes the limitations of PTAM. The ORB-SLAM can be performed by using just a monocular camera, but the scale of the map and estimated trajectory is unknown because of the unknown depth that is not observable from just one camera. Moreover, the ORB-SLAM needs to produce an initial map, as it cannot be triangulated from the very first frame. The monocular ORB-SLAM suffers from scale drift and may fail if performing pure rotations in exploration, so researchers try to use RGB-D cameras to provide the depth information to solve the above problems. KinectFusion applies the RGB-D camera to fuse all depth data from the sensor into a volumetric dense model and track the camera pose by using ICP. While the effective range of the depth data detected by Kinect is 0.4m~3.5m, the depth information of the Kinect camera is inaccurate when the surface of the scene is less than 0.4m, which makes the input point cloud of the ICP algorithm low accuracy and eventually leads to the inaccuracy of the matching result. When the geometric information of the scene is less and the point cloud is too smooth, the ICP algorithm is difficult to converge, the accuracy of the algorithm is reduced, the position and posture of the camera can not be correctly calculated, which cause the situation of "losing frame". Therefore, this system was limited to small workspaces. Recently, the ORB-SLAM2 provides a real-time SLAM library for Stereo and RGB-D cameras that computes the camera trajectory and a sparse 3D reconstruction with true scale to give us alternative ways. The optical flow method use the correspondence of the three-dimensional point and the two-dimensional pixel point of the image and tracks the two-dimensional pixels of the image through the light stream to obtain the outer parameters and track the camera.

In this paper, a combination of optical flow method and ICP is proposed. We use the ICP algorithm to reconstruct the scene and initialize the 3D point coordinates by the ICP algorithm, and then the optical flow method is used to track the location camera. When the camera tracking fails, the ICP based "lost frame retrieval" method is used to rematch. We also fuse this method and RGB-D camera ORB-SLAM to make full use of the color and depth data and refine the registration results.

2 TRACKING METHODS

2.1 ORB-SLAM2

KinectFusion of Newcombe et al. (Hong, 2015), fusing all depth data from the RGBD camera into a volumetric dense model that is used to track the camera pose using ICP, was limited to small workspaces due to its volumetric representation and the lack of loop closing. ORB-SLAM2 (Ra´ul, 2017) uses depth information to synthesize a stereo coordinate for extracted features on the image, so that the system is suitable for the input being stereo or RGB-D. This system does not need a specific structure from motion initialization as in the monocular case, because it can get the depth information from just one frame with the RGB-D cameras. The system operations are based on these features extracted from the input frame, so that the system can run whether the input image is from the stereo camera or RGB-D camera. For RGB-D cameras, the system extract ORB features on the RGB image and the system uses the same ORB features for tracking, mapping and place recognition tasks. The back end of ORB-SLAM2 is based on bundle adjustment and builds a globally consistent sparse reconstruction.

The goal of ORB-SLAM2 is long-term and globally consistent localization instead of building the most detailed dense reconstruction. Therefore, we fuse ICP Algorithm and optical flow method to present a method based on the RGB-D camera ORB-SLAM to get the dense map.

2.2 ICP Algorithm

ICP algorithm is a common algorithm for point cloud matching. The algorithm obtains the point cloud matching relationship in two coordinate systems, that is, the transformation relation of the two coordinate systems is obtained. It is necessary to find a suitable rotation matrix and the translation vector, so that the two input point clouds P and X can match (Wen, 2015), that is, to find the least square approximation coordinate transformation matrix of the two point cloud P and X. The rotation and translation matrix can be computed by quaternions. Set a rotation transformation vector as a unit quaternions $q_R = \left[q_0, q_1, q_2, q_3\right]^T$, where $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$, and obtain rotation matrix R, set translation transformation vector as $q_T = \left[q_4, q_5, q_6\right]^T$, and obtain coordinate transformation vector $q = \left[q_R \mid q_T\right]^T$, then, The problem of finding the best transformation vector for the corresponding point set can be converted to the minimization of the (1) formula.

$$f(q) = \frac{1}{N_P}\sum_{i=1}^{N_P}\|x_i - R(q_R)P_i - q_T\|^2 \quad (1)$$

2.3 The Optical Flow Method

The optical flow represents the change of the image, and it contains the information of the target movement, so we can use it to determine the motion of the target.

The basic premise of the optical flow method is as follows (Meister, 2012):

1. The luminosity is constant. The luminance of the same point will not change as time changes. This is the assumption of the basic optical flow method (all optical flow related method must be satisfied) for obtaining the basic equations of the optical flow method.

2. Small movements. The position changes are small as time changes, so that the grey level can get the partial derivative of the position. In other words, in the case of small movement, we can use the grey change caused by the change of the unit position between the front and back frames to approximate the partial derivative of the grey level to the position level, which is also an indispensable assumption of the optical flow method.

3. Uniform space. The adjacent points in a scene are adjacent points when projected onto the image, and adjacent points have uniform speed.

3 THE COMBAINATION METHOD

3.1 Advantages of Optical Flow Method and ICP Combination Algorithm

The optical flow method relies on the color image to calculate the feature points, and the coordinates of the 3D points corresponding to the 2D pixels of the initial image are required for the subsequent matching. Therefore, the optical flow method is sensitive to light and requires constant illumination. The ICP algorithm is to obtain the transformation relationship between the point cloud from the depth map and the 3D point cloud of the reconstruction scene. The matching process of the ICP method uses the depth data, and the depth map does not rely on the constant light. The ICP algorithm has a high accuracy, however, the precision of the initial depth map obtained by RGB-D camera is limited, and its effective range is 0.4m~3m, which is inaccurate when the depth is near or less than 0.4m, resulting in the inaccurate matching result of the current point cloud. In the more complex scene, the more obvious the characteristics of point cloud and the better the matching effect of the ICP point cloud matching algorithm is. In the smooth scenes, the matching point clouds are smoother, the ICP algorithm is difficult to converge, and the accuracy of the matching result is low. When the matching is inaccurate, we can see that even if the camera is fixed, the two-dimensional image after projection obviously shakes. Within the effective distance of RGB-D camera, the accuracy and stability of ICP algorithm are relatively high. By recording the matched information, it rematches more quickly when the camera tracking and localization fail. In general, the optical flow method and the ICP algorithm have their own advantages and disadvantages. In the augmented reality system, the ICP algorithm can be used to reconstruct the scene, and initialize to get the 3D point coordinates, and then the optical flow method is used to track and locate the camera. When the camera tracking fails, the "lost frame retrieval" mechanism based on the ICP algorithm is used to rematch. The combination of optical flow method and ICP algorithm can make use of depth data and color data more effectively.

3.2 The Combination Algorithm Flow

The OpenCV library function is used in the optical flow method. Image corner detection is processed at the beginning of the algorithm and the 3D point coordinates corresponding to the two-dimensional pixels are found by using the ray casting algorithm. The ray casting algorithm requires the camera external reference. We apply the position relationship of the two cameras of Kinect and the relationship between the two cameras and the world coordinate system to obtain the conversion relationship between the color camera coordinate system and the depth camera coordinate system. Thus, the depth camera external parameters are converted to the color camera external reference, and the ray casting algorithm is used to calculate the three-dimensional point coordinate corresponding to a color pixel.

The Ransac process is the process of selecting more accurate matching points from 2D points and 3D points. When the camera tracking fails, it is necessary to use the ICP algorithm to retrieve the correct camera pose. After the camera reaches steady state, the optical flow method can be used again to convert to the tracking process.

The flow chart of our method is shown in Figure1. The upper part is the pose estimation in the tracking process. The lower part shows the whole 3D reconstruction based on the ORB-SLAM. The pose estimation is the combination of the ICP and the optical flow, which makes full use of the color images and the depth maps to get an more accurate dense reconstruction.

4 EXPERIMENT

After the optical flow tracking the characteristic points, the PnP Method is used to solve the pose. When the position and posture satisfies the condition (two thresholds we set), the 3D points of the last frame are projected to the current frame. We have run the combination algorithm proposed in this paper in an Intel Core i7-6700HQ desktop computer with 8GB RAM, and evaluated the proposed algorithms on the TUM RGB-D benchmark (Sturm, 2012). In the experiment, we draw the feature points (corners) tracked by the optical flow and the re-projected feature points (re-projected corners)

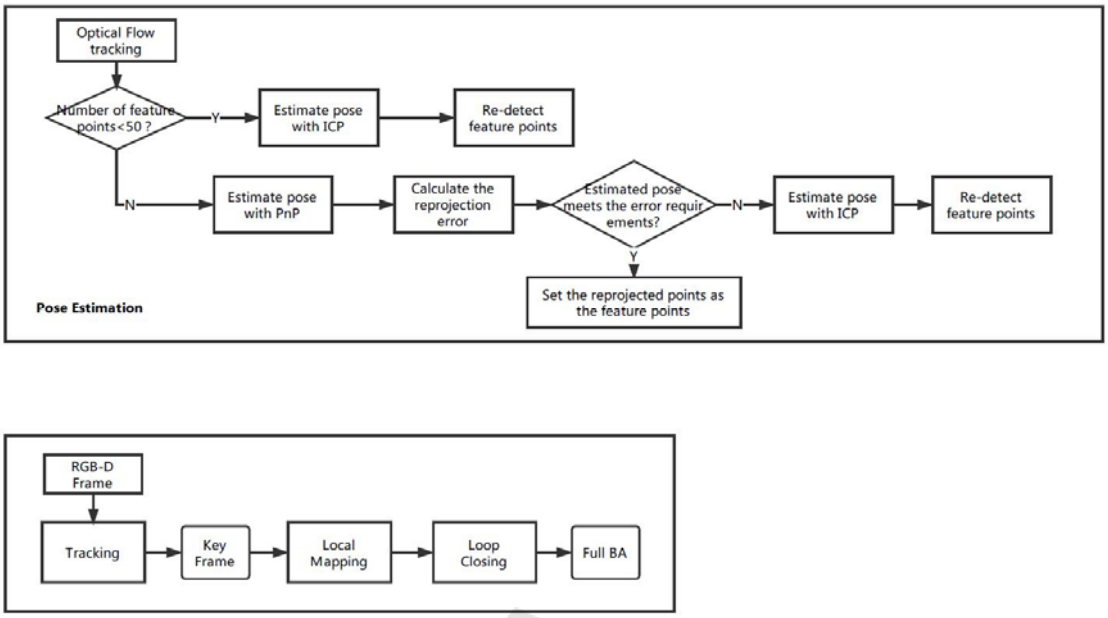

Figure1: Flow chart of the proposed combination algorithm.

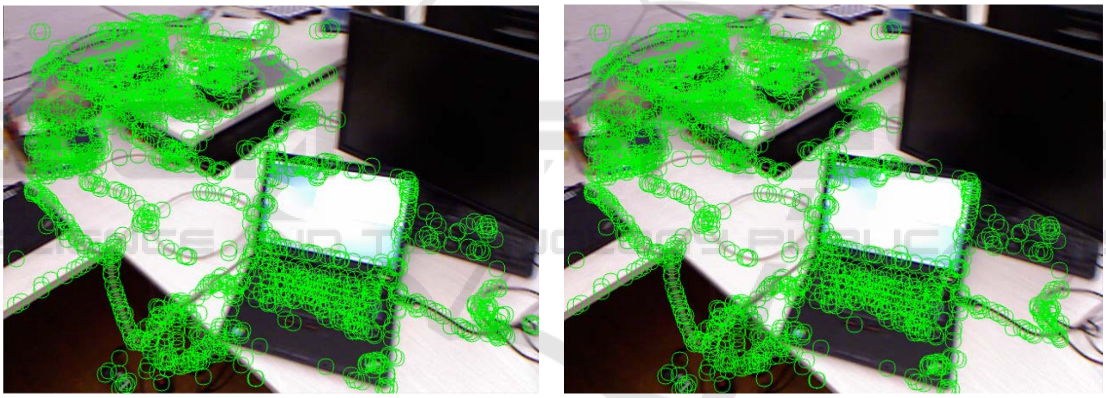

Figure 2: Feature points (left) tracked by the optical flow and the re-projected feature points (right).

respectively. The re-projection is that the 3D points of the last frame are projected again onto the pixel plane, and the calculation process does not involve the points detected from the current frame. Therefore, if the feature points (corners) and the re-projected feature points (re-projected corners) are overlapped, it shows that the pose calculation is good. From Figure2, the re-projection points are basically consistent with the tracked feature points, which indicates that the pose estimation is accurate.

5 CONCLUSIONS

In this paper, we proposed an efficient ORB-SLAM2 based registration algorithm for 3D dense reconstruction, combining the optical flow method and ICP. We use the ICP algorithm to reconstruct the scene and initialize the 3D point coordinates by the ICP algorithm, and then the optical flow method is used to track the location camera. When the camera tracking fails, we use the ICP based "lost frame retrieval" method to rematch. We also fuse this method and RGB-D camera ORB-SLAM to make full use of the color and depth data and refine the registration results. This improved ORB-SLAM based method can reconstruct dense map for practical applications. In the future research, we will add more efficient filter to reduce the point cloud repetition rate to get more accurate 3D reconstruction.

ACKNOWLEDGMENTS

This research is supported by the National Natural Science Foundation of China (NSFC) under Grant No. 61702308 and No. 51504146.

REFERENCES

1. Galvez, A., Iglesias, A., Particles warm optimization for non-uniform rational B- spline surface reconstruction from clouds of 3D data points, Inf. Sci.192(2012) 174–192, http://dx.doi.org/10.1016/j.ins.2010.11.007.
2. Hong, C., Yu, J., You, J., Chen, X., Tao, D., Multi-view ensemble manifold reg- ularization for 3D object recognition, Information Sciences, doi: 10.1016/j. ins.2015.03.032.
3. Catuhe，D., Programming with the Kinect for Windows Software Development Kit, Microsoft Press, United States, 2012.
4. Valiente，D., Gil，A., Fernández，L., Reinoso，Ó., A modified stochastic gradient descent algorithm for view-based SLAM using omni direction alimages ,Inf.Sci. 279(2014)326–337, http://dx.doi.org/10.1016/j.ins.2014.03.122.
5. Kim，D.Y., Jeon，M., Data fusion of radar and image measurements for multi- object tracking via Kalman filtering, Inf.Sci.278(2014)641–652, http://dx.doi. org/10.1016/j.ins.2014.03.080.
6. Lu，K., Wang，Q., Xue，J., Pan，W., 3D model retrieval and classification by semi- supervised learning with content-based similarity,Inf.Sci.281(2014) 703–713, http://dx.doi.org/10.1016/j.ins.2014.03.079.
7. Sun，L., Liu，Z., Sun，M.-T., Real time gaze estimation with a consumer depth camera, Information Sciences, doi: 10.1016/j.ins.2015.02.004.
8. Ala，R., Kim，D.H., Shin，S.Y., Kim，C., Park，S.-K.,A 3D-graspsynthesisalgorithmto grasp unknown objects based on graspable boundary and convex segments, Inf. Sci. 295(2015)91–106, http://dx.doi.org/10.1016/j.ins.2014.09.062.
9. G´alvez-L´opez，D. and Tard´os，J. D., “Bags of binary words for fast place recognition in image sequences,” IEEE Transactions on Robotics,vol. 28, no. 5, pp. 1188–1197, 2012.
10. Strasdat，H., Montiel，J. M. M., and Davison，A. J., “Scale drift-aware large scale monocular SLAM.” in Robotics: Science and Systems (RSS),Zaragoza, Spain, June 2010.
11. Strasdat，H., Davison，A. J., Montiel，J. M. M., andKonolige，K.,“Double window optimisation for constant time visual SLAM,” in IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain,November 2011, pp. 2352–2359.
12. Ra´ul M. A., Tard´os J. D., ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras, IEEE Transactions on Robotics, 33(5) pp.1255-1262.
13. Wen, Y. J., Hsu, C. C., Wang, W. Y.. Map building of uncertain environment based on iterative closest point algorithm on the cloud. In: Proceedings of IEEE International Conference on Mechatronics. Nagoya: IEEE, 2015: 188 ~ 190
14. Meister, S.， Izadi, S.， Kohli, P.， Hmmerle, M.， Rother, C.， Kondermann, D . , When Can We Use KinectFusion for Ground Truth Acquisition，In Workshop on Color-Depth Camera Fusion in Robotics， IEEE International Conference on Intelligent Robots and Systems，2012 .
15. Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D., A benchmark for the evaluation of rgb-d slam systems. In Intelligent Robots and Systems (IROS), IEEE/RSJ Int. Conf. on, 2012: 573–580.