OF-VO：Robust and Efficient Stereo Visual Odometry Using Points and Feature Optical Flow

ABSTRACT

Stereo visual odometry is a critical component for mobile robot navigation and safety. It estimates the ego-motion using stereo images frame by frame. In this paper, we demonstrate an approach of calculating visual odometry for indoor or outdoor robots(UGV) equipped with a stereo rig. Differ from others visual odometry, we use an improved stereo-tracking method that combines information from optical flow and stereo to estimate and control the current position of Unmanned Ground Vehicle. For feature matching, we employ the circle matching strategy in VISO-2. A high-accuracy navigation system is used to evaluate our results on challenging real-world video sequences. The experiment result indicates our approach is more accurate than other visual odometry method in accuracy and run-time.

Index Terms— Stereo vision, optical flow, circle matching, tracking

I. INTRODUCTION

In the past few decades, the area of mobile robotics and autonomous systems has attracted substantial attention from researchers all over the world, resulting in major advances and breakthroughs. Currently, for applications which mobile robots are expected to perform complicated tasks that require navigation and localization in complex and dynamic indoor and outdoor environments without any human input. As a result, the localization problem has been studied in detail and various techniques have been proposed to solve the localization problem. The simplest form of localization is to use wheel odometry methods that rely upon wheel encoders to measure the amount of rotation of robots wheels. In those methods, wheel rotation measurements are incrementally used in conjunction with the robot’s motion model to find the robot’s current location with respect to a global reference coordinate system. The wheel odometry method has some major limitations. Firstly, it is limited to wheeled ground vehicles and secondly, since the localization is incremental (based on the previous estimated location), measurement errors are accumulated over time and cause the estimated robot pose to drift from its actual location. There are a number of error sources in wheel odometry methods, the most significant being wheel slippage in uneven terrain or slippery floors.

To overcome those limitations, other localization strategies such using IMUs(Zhou Q, Lei S, Zhangguo Y U, et al. Indoor Positioning System Using Axis Alignment and Complementary IMUs for Robot Localization[J]. ROBOT, 2017, 39(3):316-323.), GPS, LASER odometry（Roy N. Stereo vision and laser odometry for autonomous helicopters in GPS-denied indoor environments[J]. Proceedings of SPIE - The International Society for Optical Engineering, 2009, 7332(1):373-375.） and most recently Visual Odometry(VO) (*Nister D, Naroditsky O, Bergen J. Visual odometry[C]// Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. IEEE, 2004:I-652-I-659 Vol.1.*)and Simultaneous Localization and Mapping(SLAM) (*Durrantwhyte H, Bailey T. Simultaneous localization and mapping: part I[J]. IEEE Robotics & Amp Amp Automation Magazine, 2006, 13(2):99 - 110.*)methods have been proposed.

The estimation of the movement of a camera, especially a stereo-camera rig, is an important task in robotics and advanced driver assistance systems. For these estimation of the movement, VO is the critical method for relative locating, especially indoors that GPS denied. VO estimates the ego-motion of robot using single or stereo cameras, which is more accurate than the wheel odometry according to Yousif et al. (2015).(Yousif, K., Bab-Hadiashar, A., & Hoseinnezhad, R. (2015). An overview to visual odometry and visual slam: applications to mobile robotics.*Intelligent Industrial Systems,1*(4), 289-311.) Pavan et al.(2018) (Pavan K U, Sahul M P V, Murthy B T V. Implementation of stereo visual odometry estimation for ground vehicles[C]// IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology. IEEE, 2018.)

The remainder of this paper is structured as follows. After the introduction, the related work is described in Section II. In Section III the proposed method is introduced. Experimental results of the proposed approach using image sequences。。。。。。。。。 are given in Section IV. Followed by the conclusion and an outlook on future work in Section V.

II. RELATED WORK

VO is a particular case of a more general problem called structure from motion (SfM). SfM simultaneously recovers relative camera poses with 3D structure of viewed object and it is commonly used in computer vision community(Scaramuzza, D., Fraundorfer, F.: Visual odometry: part I—the first 30 years and fundamentals. IEEE Robot. Autom. Mag. 18(4), 80–92 (2011)). SfM spends a lot of computation on the final refinement and global optimization of the camera pose and structure, and usually performed offline. However, the estimation of the camera poses in VO is required to be done in real-time. So far, there are many VO methods proposed which including monocular(Li R, Wang S, Long Z, et al. UnDeepVO: Monocular Visual Odometry through Unsupervised Deep Learning[J]. 2017.   Yang N, Wang R, Cremers D. Feature-based or Direct: An Evaluation of Monocular Visual Odometry[J]. IEEE Robotics & Automation Letters, 2017, PP(99):1-1.) , stereo vision(Howard A. Real-time stereo visual odometry for autonomous ground vehicles[C]// Ieee/rsj International Conference on Intelligent Robots and Systems. IEEE, 2008:3946-3952.   Pavan K U, Sahul M P V, Murthy B T V. Implementation of stereo visual odometry estimation for ground vehicles[C]// IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology. IEEE, 2018.) and RGBD methods(Schenk F, Fraundorfer F. Robust edge-based visual odometry using machine-learned edges[C]// Ieee/rsj International Conference on Intelligent Robots and Systems. IEEE, 2017.   Gutierrez-Gomez D, Mayol-Cuevas W, Guerrero J J. Dense RGB-D visual odometry using inverse depth[J]. Robotics & Autonomous Systems, 2016, 75(PB):571-583.).

A. Camera Model

This section describes the camera model used in the proposed approach. A camera model is a function which maps our 3D world onto a 2D image plane and is designed to closely model a real-world physical camera. There are many camera models of varying complexity. In this paper, we will explain the basic and most common model: the perspective camera model. Perspectives the property that objects that are far away appear smaller than closer objects, which is the case with human vision and most real world cameras. The most common model for perspective projection is the pinhole camera model which assumes that the image is formed by the intersecting light rays from the objects passing through the center of the projection with the image plane. An illustration of the perspective projection is shown in Fig. X. The pinhole perspective projection equation can be written as:

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = KX = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

where u and v are the 2D image coordinates of a 3D point with coordinates X, Y and Z, after it is projected onto the image plane, λ is a depth-scale factor, K is the intrinsic calibration matrix and contains the intrinsic parameters: f_x and f_y are the focal lengths in the x and y directions, and c_x and c_y are the 2D coordinates of the projection center.

B. Optical Flow

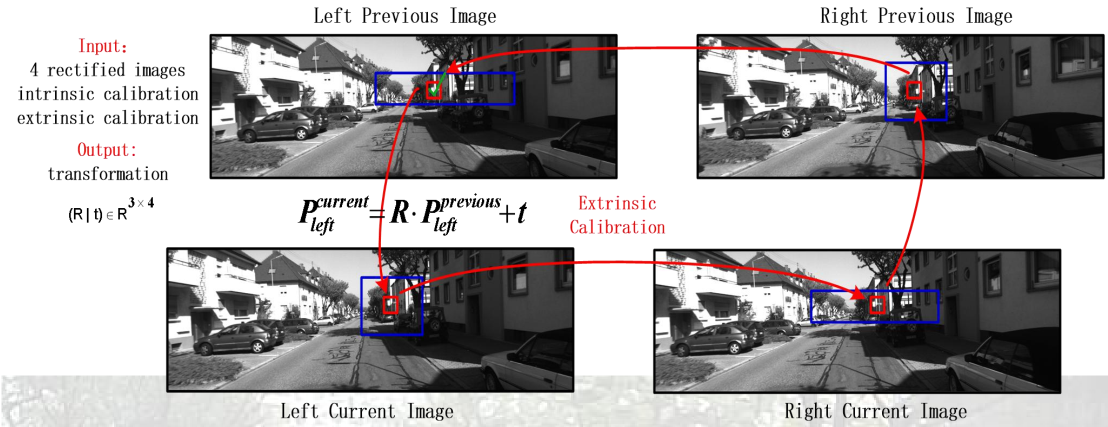
Optical flow is the motion information description of image brightness

III. VISUAL ODOMETRY PIPELINE

A. Feature Detecting and Tracking

B. Stereo Matching

C. Egomotion Estimation



双目视觉里程计系统流程如图，分别为估计线程和优化线程。

在估计线程中，为了使视觉里程计在普通 CPU 上能够达到实时要求，本文使用效率较高的 ORB 特征提取及描述算法(Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]// IEEE International Conference on Computer Vision. IEEE, 2012:2564-2571.)。通过基于金字塔的特征光流算法进行特征的匹配追踪，并使用环形匹配的策略得到相机的位姿。同时把追踪特征点数较多的帧作为关键帧，计算该帧特征点的世界坐标系的立体坐标作为地图点。

在优化线程中寻找和当前帧具有较多共同特征的关键帧作为相邻关键帧(Mur-Artal R, Tardós J D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras[J]. IEEE Transactions on Robotics, 2017, 33(5):1255-1262.)，和这些帧的地图点同时构建局部地图，并对相机位姿和匹配的地图点坐标进行局部地图的光束平差法优化(Engels C, Stewénius H, Nistér D. Bundle adjustment rules[C]// Photogrammetric Computer Vision. 2006.)。

RANSAC(RANdom SAmple Consenous) is commonly used and popular algorithm to remove outliers. The basic idea behind RANSAC is to take any two correspondences and fit line between them. Then compute distance of remaining points from this line. Consider the points which are having distance below a threshold as inliers. Repeat this step as many number if times to obtain more number of inliers.(Fraundorfer F, Scaramuzza D. Visual Odometry : Part II: Matching, Robustness, Optimization, and Applications[J]. IEEE Robotics & Automation Magazine, 2012, 19(2):78-90.)

The number of iterations N required to guarantee that a correct solution id found can be computed as.

$$N = \frac{\log(1-p)}{\log(1-(1-\varepsilon)^{S})}$$

where, p is the probability of success, ε is the percentage of outliers in the data and S is the number of data points.

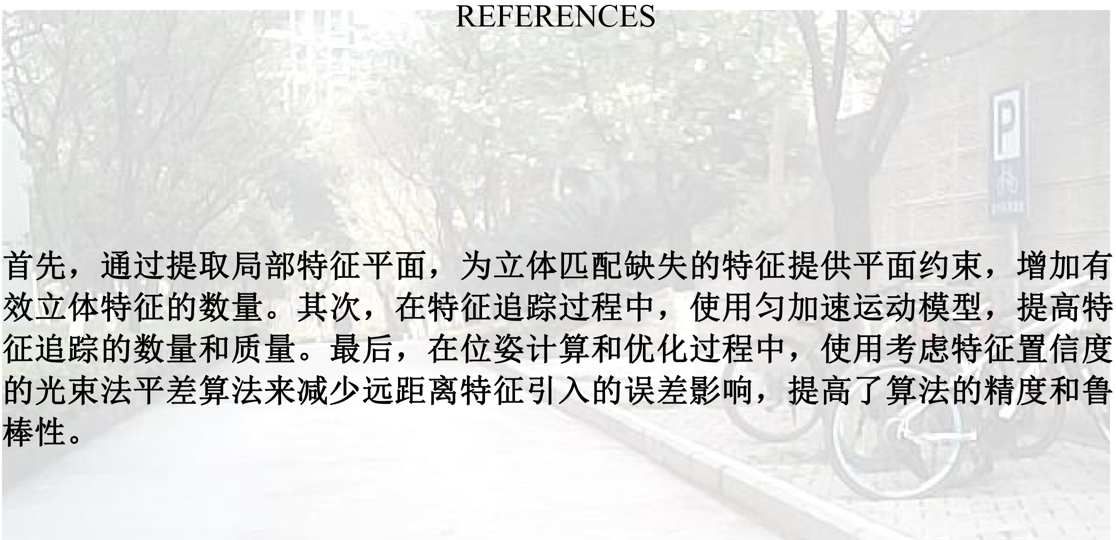
IV. EXPERIMENTAL RESULTS

The database used in this paper is the standard KITTI odometry dataset, Karlsruhe Institute of technology, Chicago. The KITTI odometry dataset comes from real outdoor environment, and include 9 stereo sequences saved in loss less PNG format.

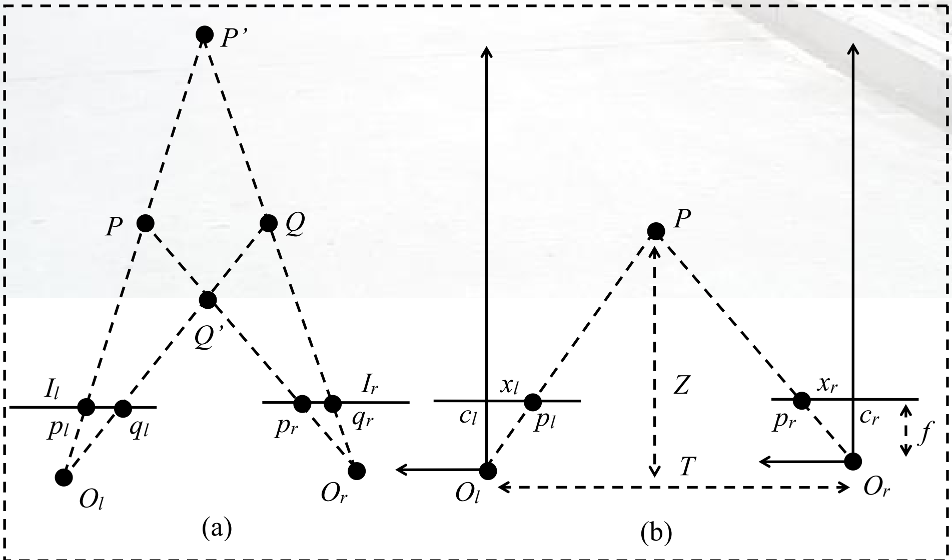
V. CONCLUSION

VI. ACKNOWLEDGEMENTS

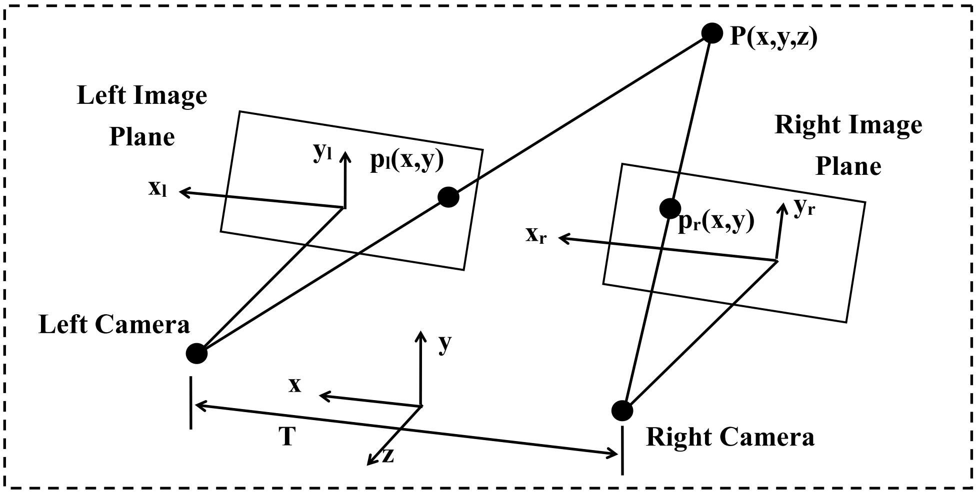
REFERENCES



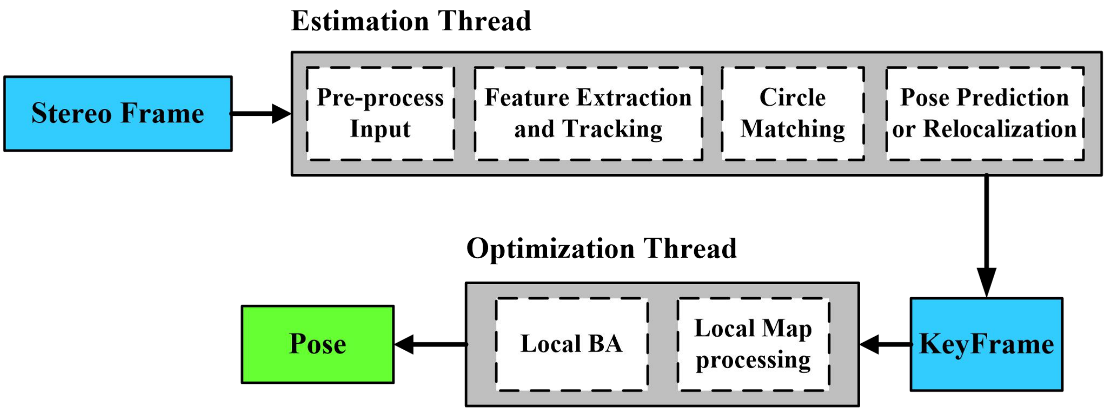
首先，通过提取局部特征平面，为立体匹配缺失的特征提供平面约束，增加有效立体特征的数量。其次，在特征追踪过程中，使用匀加速运动模型，提高特征追踪的数量和质量。最后，在位姿计算和优化过程中，使用考虑特征置信度的光束法平差算法来减少远距离特征引入的误差影响，提高了算法的精度和鲁棒性。



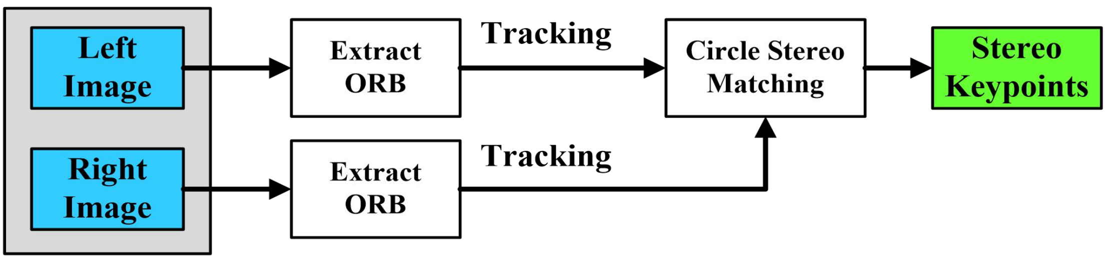
A simple stereo system



The geometry of stereo vision



(a) System Threads and Modules



(b) Input Pre-processing