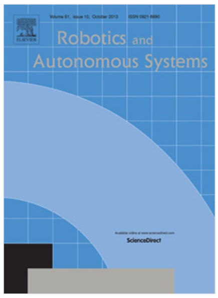
Accepted Manuscript



Robust RGB-D visual odometry based on edges and points

Erliang Yao, Hexin Zhang, Hui Xu, Haitao Song, Guoliang Zhang

PII: S0921-8890(18)30077-0
DOI: https://doi.org/10.1016/j.robot.2018.06.009
Reference: ROBOT 3048

To appear in: *Robotics and Autonomous Systems*

Received date : 31 January 2018
Revised date : 9 May 2018
Accepted date : 19 June 2018

Please cite this article as: E. Yao, H. Zhang, H. Xu, H. Song, G. Zhang, Robust RGB-D visual odometry based on edges and points, *Robotics and Autonomous Systems* (2018), https://doi.org/10.1016/j.robot.2018.06.009

Robust RGB-D Visual Odometry Based on Edges and Points

Erliang Yao[a,*], Hexin Zhang[a], Hui Xu[a], Haitao Song[a], Guoliang Zhang[b]

[a] *Department of Control Engineering , High-Tech Institute of Xi'an, Xi'an, China*

[b] *College of Controlling Engineering , Chengdu University of Information Technology, Chengdu, China*

***Abstract*—Localization in unknown environments is a fundamental requirement for robots. Egomotion estimation based on visual information is a hot research topic. However, most visual odometry (VO) or visual Simultaneous Localization and Mapping (vSLAM) approaches assume static environments. To achieve robust and precise localization in dynamic environments, we propose a novel VO based on edges and points for RGB-D cameras. In contrast to dense motion segmentation, sparse edge alignment with distance transform (DT) errors is adopted to detect the states of image areas. Features in dynamic areas are ignored in egomotion estimation with reprojection errors. Meanwhile, static weights calculated by DT errors are added to pose estimation. Furthermore, local bundle adjustment is utilized to improve the consistencies of the local map and the camera localization. The proposed approach can be implemented in real time. Experiments are implemented on the challenging sequences of the TUM RGB-D dataset. The results demonstrate that the proposed robust VO achieves more accurate and more stable localization than the state-of-the-art robust VO or SLAM approaches in dynamic environments.**

***Index Terms*—Localization, Visual Odometry, Dynamic Environments, Edge Alignment, Bundle Adjustment**

1 Introduction

In unknown or GPS-denied environments, accurate localization with visual data is one of the most active topics in Robotics. RGB-D cameras could directly provide depth information of environments, and have been used in more and more robotic applications. As a part of vSLAM (visual Simultaneous Localization and Mapping) [1], VO (Visual Odometry) [2] focuses on the egomotion estimation of consecutive images. To simplify the formulation, most SLAM and VO approaches assume static environments, namely, the information among consecutive images is consistent. However, moving objects exist everywhere, such as pedestrians, animals and cars. The image information factors that change at different times consist of the camera motion and moving objects. The moving objects destroy the static environment assumption and have negative effects on VO. Traditional VO methods obtain poor localization in dynamic environments.

Based on optimized errors, traditional VO or SLAM methods can be categorized into roughly four types. The first type is based on reprojection errors, namely, indirect methods, such as PTAM [3] and ORB-SLAM2 [4]. Point features of consecutive images are extracted and matched with invariant feature descriptors. Reprojection errors of point features are minimized for precise egomotion estimation. Due to robust point features, these methods allow large inter-frame movements, and the features can be used for loop detection to further improve the localization accuracy. The second type is based on photometric errors, namely, direct methods, e.g. DVO [5], LSD-SLAM [6] and DSO [7]. Direct methods apply intensity values of raw image data to estimate the camera motion without feature extraction. Compared with the reprojection location of features in indirect methods, direct methods adopt the intensity gradient direction and magnitude to guide the egomotion estimation. However, they assume small inter-frame motion and photometric invariance. In addition, effective loop closures cannot be integrated into direct methods. The third type is based on Euclidean space errors with the ICP (iterative closest point) algorithm, such as RGB-D SLAM [8] and KinectFusion [9]. Instead of aligning the image information as with the above two methods, this approach registers point clouds. Due to the depth uncertainty of the 3D features [10], the egomotion estimation is less accurate than that of minimizing reprojection errors. However, optimization with large sets of points and a 3D Gaussian error model could improve the performance of the egomotion