Computers and Electrical Engineering 000 (2018) 1–15

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Direct visual-inertial odometry with semi-dense mapping☆

Wenju Xu, Dongkyu Choi, Guanghui Wang*

School of Engineering, University of Kansas, Lawrence, KS 66045, USA

ARTICLE INFO

Article history:
Received 30 April 2017
Revised 1 February 2018
Accepted 1 February 2018
Available online xxx

Keywords:
Visual-inertial odometry
Sensor fusion
IMU pre-integration

ABSTRACT

The paper presents a direct visual-inertial odometry system. In particular, a tightly coupled nonlinear optimization based method is proposed by integrating the recent development in direct dense visual tracking of camera and the inertial measurement unit (IMU) pre-integration. Then a factor graph optimization is adopted to estimate the pose and position of the camera, and a semi-dense map is created simultaneously. Two sliding windows are maintained in the proposed approach. The first one, based on direct sparse odometry (DSO), is to estimate the depths of candidate points for mapping and dense visual tracking. In the second one, measurements of both the IMU pre-integration and direct dense visual tracking are fused probabilistically based on a tightly-coupled, optimization-based sensor fusion framework. As a result, the scale drift of visual odometry is compensated by the constraints from the IMU pre-integration. Evaluations on real-world benchmark datasets show that the proposed method achieves competitive results in indoor scenes.

1. Introduction

In the research area of robotics, camera motion estimation and 3D reconstruction take fundamental places in navigation and perception, such as unmanned aerial vehicle (UAV) navigation [1] and indoor reconstruction [2,3]. Among these applications, an robust camera motion tracking and 3D map of the environment are desired at the same time. Most existing methods formulate this problem as simultaneously localization and mapping (SLAM), which characterized on the sensors it used. Recent efforts include visual SLAM and visual inertial navigation system (VINS).

Visual odometry [4] estimates the depth of features, based on which, track the pose of the camera. In contrast, direct visual odometry working directly on pixels without the feature extraction pipeline is free of the issues in feature based methods. This method is able to achieve drift-free estimation for slow motion. However, direct visual odometry is also subject to failure as images can be severely blurred by changing illumination, and fast motion. Consequently, aggressive motion of the UAV [5,6] with significant large angular velocities and linear accelerations makes the state estimation subject to scale drift immediately.

Pre-integrating IMU measurements can improve visual odometry remedy this issue by providing an additional short term measurements. IMU provides noisy but outlier free measurement with high frequency, making it an ideal device for fast motion tracking. We are convinced that associating the measurements of direct visual tracking and IMU pre-integration is able to achieve robust motion estimation within challenging environment.

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. A. Chaudhary.

* Corresponding author.
E-mail address: ghwang@ku.edu (G. Wang).

https://doi.org/10.1016/j.compeleceng.2018.02.002

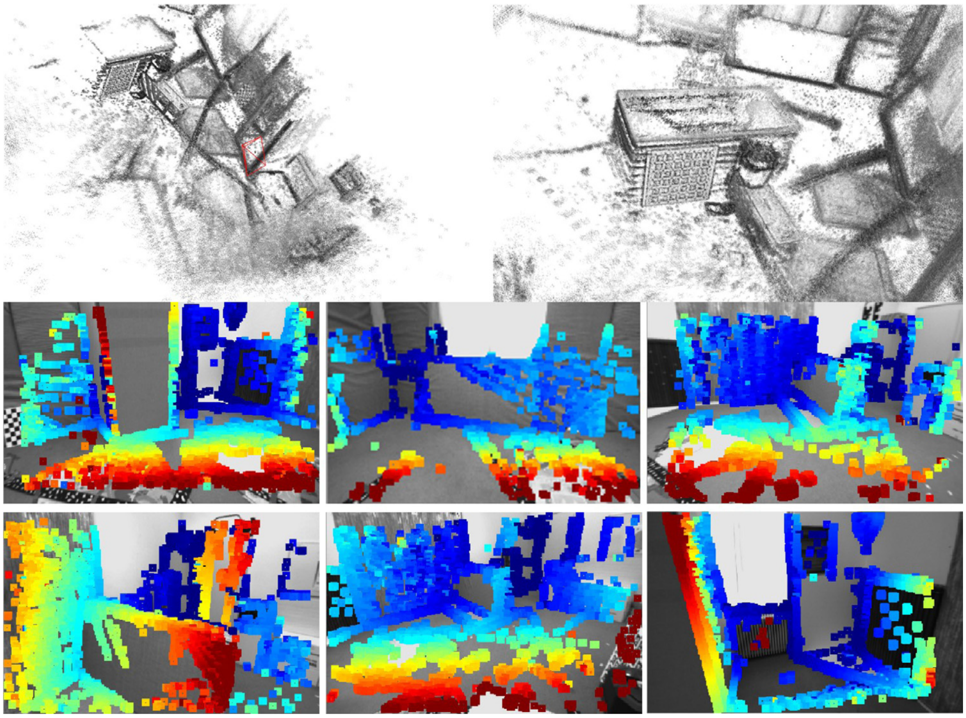

Fig. 1. 3D reconstruction and depth estimation on EuRoC dataset. The first row shows the 3D reconstruction on the V1_easy sequence and the bottom rows show the depth maps for frame tracking.

This paper works on a tightly coupled fusion system for state estimation. We combine direct dense tracking and IMU pre-integration by adopting the concept of keyframes, and make a probabilistic derivation of the IMU errors and the corresponding information matrix. This paper proposes a robust and fully integrated system for direct visual inertial odometry. The novelties of the proposed system include: 1) The combination of the direct photometric information and the edge features, which are points with sufficiently high image gradient magnitude. We work with the intensity of pixels, and the system is more robust and reliable than other methods based on detected features, as shown in Fig. 1. IMU pre-integration. The IMU pre-integration provides scale information by integrating the IMU measurements. Benefiting from the use of a factor graph, tracking and mapping are focused in a local covisible area, which has a bounded size and is independent of the global map. 3) Tightly coupled optimization. The measurements of both the IMU pre-integration and the dense visual tracking are fused probabilistically within a single tightly-coupled, optimization-based framework. In this paper, the dense visual tracking results provide the visual constraints between current frame and the reference frame, while the IMU pre-integration provides constraints between the two consecutive frames.

In the remainder of the paper, we first review some related state-of-the-art works in Section 2, followed by the formulation of the problem in Section 3. In Section 4, an overview of the system is described, followed by the details of the IMU and visual measurement. The semi-dense mapping is introduced in Section 5. Section 6 shows implement details. The experimental results and evaluations are discussed in Section 7. At last, we make a conclusion for the paper in Section 8.

2. Related work

Simultaneously localization and mapping (SLAM) has a long history in monocular scenarios, prominent examples include dense tracking and mapping (DTAM) [4], semi-direct visual odometry (SVO) [7] and large-scale direct (LSD) monocular SLAM [8], which work on sparse features and estimate the camera motion through a prediction and correction fashion. Direct methods work directly on image intensities and attract a lot of attention since they are robust and computational efficient without feature detection. It is demonstrated that direct methods are more suitable for dense mapping than feature based method, and when enhanced by edge alignment [9], they can deal with changing illumination and fast motion. More recently, direct sparse odometry (DSO) [3] presented an impressive semi-dense 3D reconstruction of the environment through a sliding window optimization. This direct method minimizes the photometric error of points with sufficiently high image gradient magnitude (edges) using a non-linear optimization framework. In addition, another type of approach is a batch factorization based structure from motion (SfM) algorithm [10–12].

More recently, people pay great attention to the 3D SLAM,