Robust Large Scale Monocular Visual SLAM

Guillaume Bourmaud, Rémi Mégret
Univ. Bordeaux, CNRS, IMS, UMR 5218, F-33400 Talence, France

1 Introduction

Estimating a 3D model of the environment in which a camera evolves as well as its trajectory, also known as Visual Simultaneous Localization And Mapping (VSLAM), is an important problem for the computer vision community. Indeed, a large number of applications, such as image-based localization or augmented reality, assume that a 3D model of the environment has been previously reconstructed. Thus, being able to accurately estimate this 3D model is essential in order for these applications to operate correctly.

In monocular VSLAM, one of the major difficulties, compared to stereo VSLAM, consists in the fact that the scale of the scene is not observed. In order to prevent scale drift, loop closures (i.e when the camera comes back at a place already visited) need to be detected. However, a large environment usually contains places that look alike. Thus when a camera evolves in such an environment, wrong loop closures may be detected, resulting in an erroneous 3D model.

2 Contribution

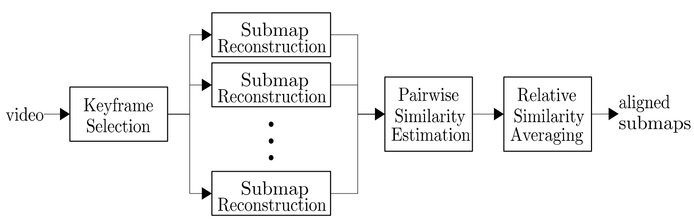

Figure 1: Our Monocular VSLAM Framework

We propose a novel robust monocular VSLAM algorithm which is able to operate on long challenging videos. It consists in 4 modules (Keyframe Selection, Submap Reconstruction, Pairwise Similarity Estimation and Relative Similarity Averaging) arranged as illustrated in Fig.1. First of all, submaps (parts of the camera trajectory and the unknown environment) are robustly and accurately estimated using the so-called *Known Rotation Problem* [7]. We then build a graph of relative 3D similarities (computed between the submaps). In order to reject the outlier relative 3D similarities coming from wrong loop closures, we propose a simple and efficient *outlier removal algorithm*. Finally, to obtain a scalable monocular VSLAM framework, we derive a *loopy belief propagation* algorithm which is able to align a large number of submaps very efficiently.

The contribution of this paper is threefold:

1. A novel visual odometry approach based on the so-called *Known Rotation Problem* that allows to robustly estimate each submap independently.
2. A simple and efficient *outlier removal algorithm* to reject the outlier relative 3D similarities coming from wrong loop closures.
3. A *loopy belief propagation* algorithm which is able to align a large number of submaps very efficiently.

3 Results

The method has been validated experimentally and compared to the two most recent state of the art algorithms which it outperforms both qualitatively and quantitatively (see Fig.2 and Fig.3). Moreover, in all our experiments (4 different cameras with different resolutions), the parameters of our method have been set once and for all proving the flexibility of the proposed approach.

	Proposed	[2]	[3]	[5]	[4]	[1]
Uses Depth	No	No	No	No	Yes	Yes
fr2/desk	**2.22**	4.52	13.50	x	**1.77**	9.5
fr2/xyz	**1.28**	1.47	3.79	24.28	**1.18**	2.6

Figure 2: Quantitative comparison on the TUM RGB-D dataset. The figures represent the absolute trajectory RMSE (cm) [8]. The results of [1, 2, 3, 4, 5] are taken from Fig.9 in [2].

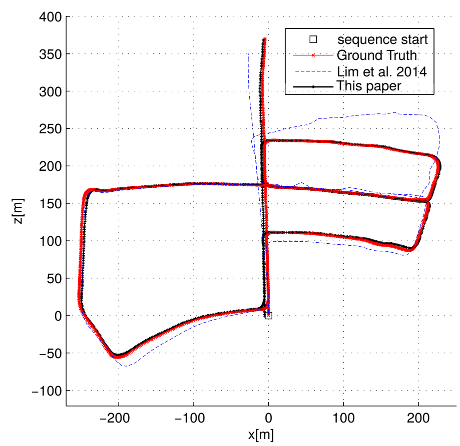

Figure 3: Qualitative comparison on the camera trajectories estimated with the approach proposed in this paper and [6] on sequence 05 of the KITTI dataset. Most of the time, the camera trajectory estimated with our approach overlaps with the ground truth as opposed to [6] which deviates from the real trajectory.

[1] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard. An evaluation of the RGB-D SLAM system. In *ICRA*, pages 1691–1696, 2012.

[2] J. Engel, T. Schops, and D. Cremers. LSD-SLAM: Large-scale direct monocular SLAM. *ECCV*, pages 834–849, 2014.

[3] J. Engel, J. Sturm, and D. Cremers. Semi-dense visual odometry for a monocular camera. In *ICCV*, pages 1449–1456, 2013.

[4] C. Kerl, J. Sturm, and D. Cremers. Dense visual SLAM for RGB-D cameras. In *IROS*, pages 2100–2106, 2013.

[5] G. Klein and D. Murray. Parallel tracking and mapping for small ar workspaces. In *ISMAR*, 2007.

[6] H. Lim, J. Lim, and H. J. Kim. Real-time 6-dof monocular visual SLAM in a large-scale environment. In *ICRA*, 2014.

[7] C. Olsson, A. Eriksson, and R. Hartley. Outlier removal using duality. In *CVPR*, pages 1450–1457, 2010.

[8] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A benchmark for the evaluation of RGB-D SLAM systems. In *IROS*, pages 573–580. IEEE, 2012.

Acknowledgments The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007- 2013) under grant agreement 288199 - Dem@Care.

The full paper is available at the Computer Vision Foundation webpage.