VITAMIN-E: VIsual Tracking And MappINg with Extremely Dense Feature Points – Supplementary Material –

Masashi Yokozuka, Shuji Oishi, Thompson Simon, Atsuhiko Banno
Robot Innovation Research Center,
National Institute of Advanced Industrial Science and Technology (AIST), Japan
{yokotsuka-masashi, shuji.oishi, simon.thompson, atsuhiko.banno}@aist.go.jp

This supplementary material provides additional experimental results and the detail of optimization method.

1. Additional Experiments

We conducted additional experiments on the TUM RGB-D benchmark[3] dataset and ICL-NUIM[2] noisy synthetic dataset. The results are shown in Table 1, which cites scores of competitors from [1] under the same conditions. The parameters of VITAMIN-E were fixed in all experiments.

Our method worked well and achieved a relatively high trajectory accuracy in a variety of scenes. The worst case was Office Room 1, in which the average trajectory error of VITAMIN-E was 40.1 cm: the camera explored a textureless environment, where only edges of the floor and walls could be seen at 18 sec, which led to difficulty in extracting curvature extrema and thus unstable tracking. Edge-based SLAMs, such as LSD-SLAM, achieved better localization performance in such scenes, whereas for point-based SLAMs' (including our method), performance tended to degrade.

2. Optimization

Alternating optimization (AO) is used in some contexts such as color map optimization [5]. In structure from motion (SfM), AO is called "resection intersection (RI)," which optimizes camera poses and feature positions alternately. [4] concluded that although RI has numerical stability in optimization, it is not suitable for SfM because of its slow convergence compared with the standard Gauss–Newton method which optimizes all variables simultaneously.

Table 1. Trajectory errors in the additional datasets. The results of other methods except our method were obtained from [1].

		Our method	DSO	SVO	ORB-SLAM (with loop closure)	ORB-SLAM (w/o loop closure)	LSD-SLAM (with loop closure)	LSD-SLAM (w/o loop closure)
TUM RGB-D	fr2_desk	1.7 cm	-	6.7 cm	**0.9** cm	-	4.5 cm	-
	fr2_xyz	0.4 cm	-	0.8 cm	**0.3** cm	-	1.5 cm	-
ICL-NUIM	Living Room 0	**0.9** cm	1.0 cm	2.0 cm	-	1.0 cm	-	12.0 cm
	Living Room 1	11.0 cm	**2.0** cm	7.0 cm	-	2.0 cm	-	5.0 cm
	Living Room 2	3.1 cm	6.0 cm	10.0 cm	-	7.0 cm	-	**3.0** cm
	Living Room 3	**2.4** cm	3.0 cm	7.0 cm	-	3.0 cm	-	12.0 cm
ICL-NUIM	Office Room 0	31.6 cm	21.0 cm	34.0 cm	-	**20.0** cm	-	26.0 cm
	Office Room 1	40.1 cm	83.0 cm	28.0 cm	-	89.0 cm	-	**8.0** cm
	Office Room 2	**3.8** cm	36.0 cm	14.0 cm	-	30.0 cm	-	31.0 cm
	Office Room 3	**5.5** cm	64.0 cm	8.0 cm	-	64.0 cm	-	56.0 cm

However, we found that this is not the case in visual odometry (VO). VO sequentially optimizes variables frame by frame; thus, most old variables are already optimized and we only need to optimize the latest camera pose and feature points. Whereas the Gauss–Newton method optimizes all variables, the proposed subspace Gauss–Newton method inspired by RI efficiently updates partial variables. Specifically, our method enables RI to eliminate some variables by leveraging the Schur complement. Variable elimination plays an important role in VO because the number of variables increases with each frame. Previous RI only solves nonlinear optimization by partially updating variables fixing the other, which results in high computational costs. By constrast, the proposed subspace Gauss–Newton method applies RI to nonlinear optimization problems after quadratic approximation, that is, $Ax{=}b$, using the Schur-complement, thus enabling the real-time RI optimization required in Visual SLAM.

References

[1] Christian Forster, Zichao Zhang, Michael Gassner, Manuel Werlberger, and Davide Scaramuzza. SVO: Semidirect visual odometry for monocular and multicamera systems. *IEEE Trans. on Robotics*, 33(2):249–265, 2017.

[2] Ankur Handa, Thomas Whelan, John McDonald, and Andrew J. Davison. A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM. In *IEEE Intl. Conf. on Robotics and Automation, ICRA*, Hong Kong, China, May 2014.

[3] Jürgen Sturm, Nikolas Engelhard, Felix Endres, Wolfram Burgard, and Daniel Cremers. A benchmark for the evaluation of rgb-d slam systems. In *Proc. of the International Conference on Intelligent Robot Systems (IROS)*, Oct. 2012.

[4] Bill Triggs et al. Bundle adjustment - a modern synthesis. In *Proc. of the Int. WS on Vision Algorithms: Theory and Practice*, 2000.

[5] Qian-Yi Zhou et al. Color map optimization for 3d reconstruction with consumer depth cameras. *ACM Trans. Graph.*, 33(4):155:1–155:10, 2014.