Imaging and Applied Optics 2018 (3D, AIO, AO, COSI, DH, IS, LACSEA, LSC, MATH, pcAOP) © OSA 2018

Monocular SLAM Using Probabilistic Combination of Point and Line Features

Yang Li, Bing Yu, Chao Ping Chen*, Nizamuddin Maitlo, Wenbo Zhang, and Lantian Mi
Smart Display Lab, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China
E-mail: ccp@sjtu.edu.cn

Abstract: We propose a monocular SLAM, in which a probabilistic combination of point and line features is adopted. Compared with point-based ORB-SLAMs, our solution is effective in improving both the accuracy and robustness under low-textured scenarios.
OCIS codes: (150.6910) Three-dimensional sensing; (150.0150) Machine vision; (150.1135) Algorithms.

1. Introduction

Last decade witnessed a boom of the simultaneous localization and mapping (SLAM) in the applications of augmented reality [1], autonomous vehicles, etc. Following a groundbreaking work made by Klein and Murray in 2007 [2], a variety of real-time SLAMs have been proposed and studied. Among them, the most successful one is the point-based SLAMs, which are robust against image noise and large geometric distortions. However, this type of SLAM can easily fail under low-textured scenarios due to the lack of key points. Motivated by this issue, a monocular SLAM, in which both the points and lines are exploited through a probabilistic combination, is introduced, as shown in Fig. 1.

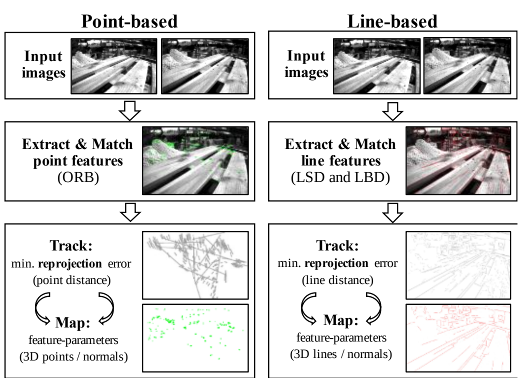


Fig. 1. Features of points and lines.

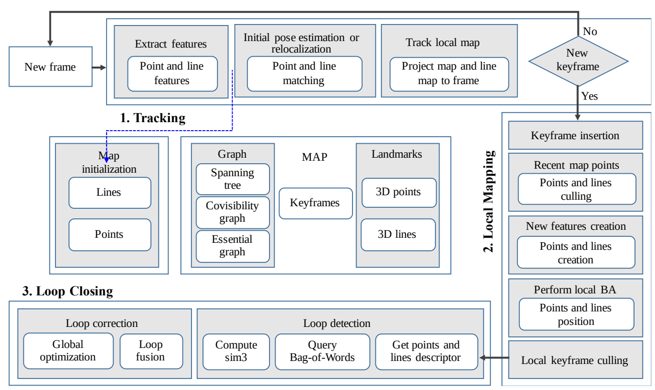


Fig. 2. Framework of the proposed SLAM.

2. Proposed solution

Fig. 2 outlines the framework of the proposed SLAM, which mainly consists of three threads that run in parallel. Thread 1 TRACKING: to localize the camera according to the features—i.e. points and lines—which are matched to the local map and to minimize the re-projection error. Thread 2 LOCAL MAPPING: to manage and optimize the local maps. Thread 3 LOOP CLOSING: to detect loops and correct the accumulated drift. Finally, both structure and motion will be optimized via full bundle adjustment.

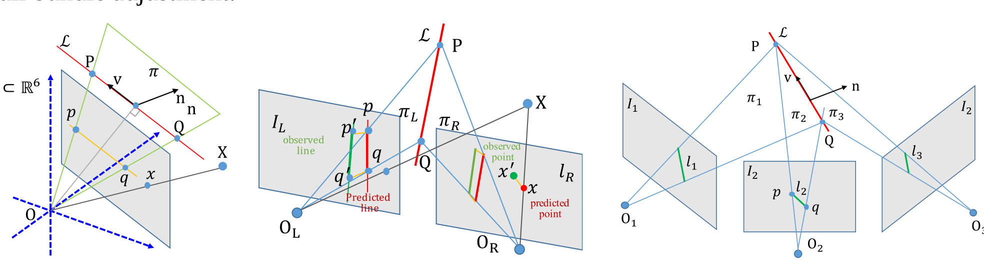
$$\mathcal{L} = (\mathrm{n}^T, \mathrm{v}^T)^T$$

$$\mathcal{L} = \begin{bmatrix}\tilde{X}_1 \times \tilde{X}_2 \\ r_2\tilde{X}_1 - r_1\tilde{X}_2\end{bmatrix} = \begin{bmatrix}\mathrm{n} \\ \mathrm{v}\end{bmatrix} \in \mathrm{P}^5 \subset \mathbb{R}^6$$

$$\mathrm{U} = \begin{bmatrix}\frac{n}{||n||} & \frac{v}{||v||} & \frac{n \times v}{||n \times v||}\end{bmatrix}$$

$$\mathrm{W} = \begin{bmatrix}\omega_1 & -\omega_2 \\ \omega_2 & \omega_1\end{bmatrix}$$

$$(\mathrm{W}, \mathrm{U}) \in SO(3) \times SO(2)$$



(a) (b) (c)

Fig. 3. (a) representations of points and lines, (b) re-projection errors, and (c) camera initialization.

Compared to points, the representation of lines is far more complex. In order to reduce this complexity, two approaches—including Plücker line coordinate $\mathcal{L}$ and orthonormal representation (U, W) by QR decomposition [3]—are employed, as shown in Fig. 3(a). The former is for initialization of observed lines and projection of 3D lines, whereas the latter for graph optimization. First, lines are detected by LSD [4], and then matched in pairs by LBD [5].

Second, re-projection errors of points and lines, see Fig. 3(b), are combined by a probabilistic weight, which can be obtained from their covariance matrices and propagation of Gaussian distribution errors. Third, a Jacobians matrix is calculated during the pose graph optimization. Fourth, three consecutive images are used to estimate an approximate initial map, see Fig. 2 (c). Besides, loop closing is computed with the bag-of-words model, taking advantage of the descriptive ability of the combined point and line features.

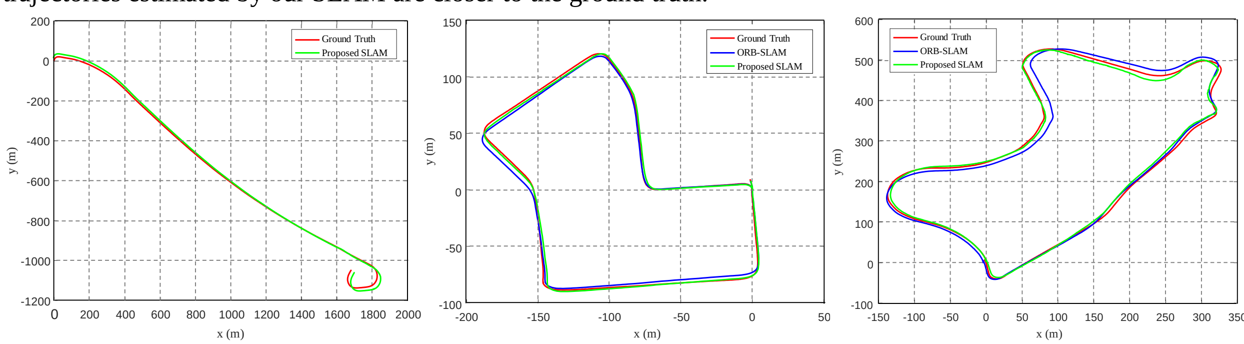
3. Experimental results

Our experiments are carried out on a PC platform (CPU: Intel Core i7-6600@2.60 GHz, Memory: 16 GB RAM, OS: 64-bit Linux Ubuntu). On an indoor dataset, known as TUM RGB-D benchmark, localization accuracies of our SLAM and point-based ORB-SLAM [6] are compared. For the sake of decreasing the amount of calculation, only the keyframes, which are extracted from the image sequence, are calculated. Table 1 summarizes the absolute median RMS errors of the camera trajectories over five executions in each sequence. For all 6 different indoor scenes, our SLAM exhibits a higher accuracy in camera trajectory than that of ORB-SLAM.

Table 1. Localization accuracies of ORB-SLAM and the proposed SLAM

TUM RGB-D sequences	Absolute median RMS errors of camera trajectory (cm)	
	ORB-SLAM	Proposed SLAM
fr1/xyz	1.38	**1.34**
fr1/desk	1.69	**1.66**
fr1/floor	5.64	**3.43**
fr2/xyz	0.64	**0.56**
fr2/360_kidnap	4.99	**3.78**
fr3/office	4.35	**2.13**

On an outdoor dataset, namely KITTI, where sequences of 01 (highway), 07 (urban) and 09 (urban) are selected, trajectories estimated by our SLAM and ORB-SLAM are compared with reference to the ground truth, as shown in Fig. 4. For sequence 01, a typical low-textured scenario, our SLAM works properly while ORB-SLAM fails, meaning a better robustness. For sequences 07 and 09, both SLAMs can work by successfully closing the loops, but the trajectories estimated by our SLAM are closer to the ground truth.



(a) (b) (c)

Fig. 4. Estimated trajectories of our SLAM and ORB-SLAM with respect to the ground truth at sequences of (a) 01, (b) 07, and (c) 09.

4. Conclusions

A monocular SLAM, based on both points and lines, is proposed. The overall performance, regarding the localization, trajectory estimation and loop closing, is tested on TUM RGB-D benchmark and KITTI dataset, respectively. For low-textured scenarios, the proposed SLAM is eligible in improving both the accuracy and robustness.

5. References

[1] L. Zhou, C. P. Chen, Y. Wu, Z. Zhang, K. Wang, B. Yu, and Y. Li, "See-through near-eye displays enabling vision correction," Opt. Express **25**(3), 2130-2142 (2017).

[2] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in 6th International Symposium on Mixed and Augmented Reality, (IEEE and ACM, Nara, Japan, 2007), pp. 250-259.

[3] G. Zhang, J. H. Lee, J. Lim, and I. H. Suh, "Building a 3-D line-based map using stereo SLAM," IEEE Trans. Robot. **31**(6), 1364-1377 (2015).

[4] R. G. von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall. "LSD: a fast line segment detector with a false detection control," IEEE Trans. Pattern Anal. Mach. Intell. **32**(4), pp. 722-732 (2010).

[5] L. Zhang and R. Koch, "An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency," J. Vis. Commun. Image. R. **24**(7), 794-805 (2013).

[6] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," IEEE Trans. Robot. **31**(5), 1147-1163 (2015).