

Line Feature, Vanishing Points and Manhattan World Analysis

Subjects: Automation & Control Systems

Contributor: Xiaojing He, Baoquan Li, Shulei Qiu, Kexin Liu

In conventional point-line visual inertial odometry systems under indoor environment, consideration of spatial position recovery and line feature classification can improve localization accuracy. A monocular visual inertial odometry based on structured and unstructured line features of vanishing points is proposed. First, degeneracy phenomenon caused by special geometric relationship between epipoles and line features is analyzed in process of triangulation, and a degeneracy detection strategy is designed to determine the location of the epipoles. Then, considering the property that vanishing point and epipole coincide at infinity, vanishing point feature is introduced to solve the degeneracy and direction vector optimization problem of line features. Finally, threshold constraints are used to categorize straight lines into structural and non-structural features under the Manhattan world assumption, and the vanishing point measurement model is added to the sliding window for joint optimization. Comparative tests on the EuRoC and TUM-VI public datasets validated the effectiveness of the proposed method.

Keywords: simultaneous localization and mapping ; vanishing points ; structural line ; indoor positioning

1. Introduction

With the rapid development of unmanned aerial vehicles (UAVs), self-driving vehicles, and mobile robots, the requirements for positioning technology are increasing ^[1]. While traditional simultaneous localization and mapping (SLAM) methods are low-precision, perturbation-prone, and poor in real time ^{[2][3]}, the visual-inertial odometer (VIO) method realizes navigation of the carrier by combining measurement data from inertial sensors and visual information from camera images ^[4], thus basically solving the shortcomings of technologies such as GPS ^[5]. Among them, VIO fuses a monocular camera with an IMU as a sensor to jointly realize the pose estimation of the moving body and the reconstruction of the 3D environment ^{[6][7]}. Monocular vision is categorized into feature and direct methods based on front-end visual odometry ^[8]. The feature-point method calculates camera pose with respect to map points by minimizing the reprojection error, whereas the direct method minimizes photometric error and uses pixel gradients for semi-dense or dense map building in feature-deficient settings ^[9]. Filter-based and optimization-based approaches in state estimation for dynamic systems with given noise measurements dominate the field of VIO ^[10]. Meanwhile, initialization methods of VIO are divided into two categories: loosely coupled ^[11], where the results are fused after the IMU and camera perform their own motion estimation separately ^[12], and tightly coupled, where state estimation is performed after the IMU and camera jointly construct the equations of motion ^[13]. Tight coupling can add new states to the system vector according to the characteristics of each sensor, which better utilizes the complementary nature of various types of sensors ^[14].

In VIO algorithms, point features have the advantage of being easy to detect and track, and are often used as the basic unit of pose estimation in many traditional methods ^[13]. However, point features are susceptible to environmental complexity, illumination, and camera motion speed ^[14], and cannot systematically meet the robustness requirements in some challenging scenarios ^[15]. Straight lines carry a lot of scene information and provide more constraints for the optimization process ^[16]; therefore, line features with high dimensional characteristics have often been used as compensation for point feature extraction in recent years ^[17]. PL-VIO ^[18] integrates the measurement of line features into VIO and successfully adds the error term of line feature construction by designing two representation methods of line features. However, the tracking and matching of line features varies widely based on different algorithms. PL-VINS ^[19] and PLF-VINS ^[20] optimize and complement the extraction of features to achieve a high degree of line feature reduction. PL-SVO ^[21] adds equidistant sampling of line feature endpoints to the semi-direct monocular visual, which avoids feature matching steps by virtue of the advantage of the direct method, effectively shortens the processing time of line features, and improves the stable pose estimation in semantic environments. However, since the line features are arranged and combined with countless point features, a lot of errors will inevitably occur in many point feature-based processing methods, thus affecting the localization performance of the system. UV-SLAM ^[22] proposes vanishing point features,

which are obtained from line features under the premise that vanishing point measurements are the only mapping solution, thus solving geometrical problems and improving the robustness of the system.

2. Line Feature Analysis

In the VIO study, line features had higher stability in harsh environments. The PL-SLAM ^[23] system proposed by Pumarola et al. adds a line segment detector (LSD) ^[24] on top of ORB-SLAM ^[25], which implements point feature and line feature extraction in SLAM. Fu et al. achieved real-time extraction by modifying the hidden parameters of the LSD algorithm ^[19]. Lee et al. fused corner point and line features to construct point-line residuals by utilizing the similarity of their positions, and the two features complemented each other to enrich the structural information of the carried environment ^[20]. In the above study, LSD was widely used as a line feature extractor, but it still failed to be accurate in scenarios with noise. Akinlar et al. proposed an EDLines linear timeline segment detector ^[26], which utilizes the edge drawing (ED) algorithm to detect edge pixels continuously, ultimately improving the detection speed while controlling the number of false detections. Liu et al. proposed a line segment extraction and merging algorithm to achieve feature tracking between points and lines with geometric constraints based on the removal of short and useless line segments ^[27]. Suárez et al. proposed ELSESED, which utilizes a local line segment growth algorithm to align pixels by connecting gradients at breakpoints, greatly improving the efficiency of line segment detection ^[28]. The PLI-VINS proposed by Zhao et al. designs adaptive thresholds for line segment extraction algorithm and constructs an adjacency matrix to obtain the orientation of the line segments, which greatly improves the efficiency of the line feature processing in indoor point and line vision SLAM systems ^[29]. Bevilacqua et al. proposed probabilistic clustering methods using descriptors based on grid computation, which reduces the number of samples within the grid through a special pixel selection process to design new image representations for hyperspectral images ^[30]. The analysis of line features is summarized in **Table 1**.

Table 1. Summary of line feature analysis.

Line Feature Analysis	Characteristics
^[23]	Combining line features with point features using the LSD algorithm.
^[19]	Modifying the parameters of the LSD algorithm for real-time extraction of line features.
^[20]	Fusing line features and corner features using positional similarity.
^[26]	Detection of line features using the ED algorithm to improve extraction speed.
^[27]	An algorithm capable of extracting line features with geometric constraints is proposed.
^[28]	Improving line segment detection efficiency with ELSESED with a local line segment growth algorithm.
^[29]	Setting an adaptive threshold algorithm for the line segment extraction process.
^[30]	Proposed probabilistic clustering of line segments using descriptors based on grid computing.

3. Vanishing Points Analysis

Geometrically, a vanishing point on a spatial line is formed by intersecting the image plane with a ray parallel to the line and passing through the center of the camera ^[31]. Lee et al. searched for line features based on vanishing points before fusing parallel 3D straight lines and constructed parallel line residuals by removing anomalies during the clustering of multi-view lines ^[20]. In the camera calibration method ^[32] proposed by Chuang et al., multiple vanishing points are utilized to estimate the camera's internal parameters, simplifying the calculations associated with calibration. Liu et al. proposed a two-module approach for estimating vanishing points, where the two modules are used to track clustered line features and generate a rich set of candidate points, respectively ^[33]. Camposec et al. detected vanishing points based on the inertial measurement and constantly updated it by the least squares method, which provided a global rotation observation around the gravity axis, thus correcting the yaw angle drift in pose estimation ^[34]. Li et al. projected the direction of the vanishing point of the 3D line and plane normal onto the tangent plane of the unit sphere, which estimates the rotation matrix of each image frame ^[35]. Li et al. involved the vanishing point in a deep multi-task learning algorithm, which was used to realize the adjustment of the camera, and then went through three sub-tasks to improve the detection accuracy of the vanishing point, and finally followed the track segmentation to determine the perimeter of the intruded area ^[36]. The analysis of vanishing points is summarized in **Table 2**.

Table 2. Summary of vanishing points analysis.

Vanishing Points Analysis	Characteristics
[20]	Proposing parallel line features based on clustering of vanishing points.
[32]	Estimating the intrinsic parameters of the camera using vanishing points for camera calibration.
[33]	Proposing a two-module method for estimating vanishing points to ensure precise vanishing points for each frame's cluster of line features.
[34]	Inertial-based measurements to detect vanishing points and continuous updating by the least squares method.
[35]	Involvement of vanishing points in deep multitask learning algorithms for camera tuning.
[36]	Projecting the vanishing point directions of 3D lines and plane normals onto the tangent plane of a unit sphere for estimating the rotation matrix.

4. Manhattan World Analysis

In the Manhattan world hypothesis, all planes in the world are aligned with three main directions, and utilizing the constraint relationship of the three axes can improve the building effect of the map and participation of line segments in SLAM. Kim et al. used structural features to estimate rotational and translational motions in RGB-D images, which were used to reduce drift during motion [37]. Li et al. first utilized a two-parameter representation of line features, using Manhattan world structure lines as features of the mapping, which not only encodes the global orientation information but also eliminates the cumulative drift of the camera [38]. StructVIO, proposed by Zou et al., uses the Atlanta world model, which contains the Manhattan world with different orientations, defines the starting frame coordinate system to represent each heading direction, and is constantly corrected by the state estimator as a state variable [39]. Lu et al. redefined the classification of line features based on the advantage that unstructured lines are not constrained by the environment and proposed frame-to-frame and frame-to-map line matching methods to initialize the two types of line segments, which significantly improved the trajectory accuracy of the system [40]. Peng et al. proposed a new Manhattan frame (MF) re-identification method based on the rotational constraints of MF-matched pairs, and a spatiotemporal consistency validation method to further optimize the global beam method parity energy function [41]. The analysis of the Manhattan world assumption is summarized in **Table 3**.

Table 3. Summary of Manhattan world analysis.

Manhattan World Analysis	Characteristics
[37]	Utilizing structural features in RGB-D images to estimate rotational and translational motion, reducing drift during motion.
[38]	Utilizing structured lines in the Manhattan world as mapping features to encode global directional information.
[39]	Utilizing the Atlanta world model containing different Manhattan directions to define a starting frame coordinate system representing each heading direction.
[40]	Redefining the classification of line features is based on the advantage that unstructured lines are not constrained by the environment.
[41]	Proposing a new Manhattan frame re-identification method based on rotation constraints from MF match pairs, improving the accuracy of vanishing point estimation.
[36]	Projecting the vanishing point directions of 3D lines and plane normals onto the tangent plane of a unit sphere for estimating the rotation matrix.

References

1. Jeon, J.; Jung, S.; Lee, E.; Choi, D.; Myung, H. Run your visual-inertial odometry on NVIDIA jetson: Benchmark tests on a micro aerial vehicle. *IEEE Robot. Autom. Lett.* 2021, 6, 5332–5339.
2. Campos, C.; Elvira, R. ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and Multi-map SLAM. *IEEE Trans. Robot.* 2021, 37, 1874–1890.

3. Yang, N.; Stumberg, L.; Wang, R.; Cremers, D. D3VO: Deep depth, deep pose and deep uncertainty for monocular visual odometry. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
4. Teng, Z.; Han, B.; Cao, J.; Hao, Q.; Tang, X.; Li, Z. PLI-SLAM: A Tightly-Coupled Stereo Visual-Inertial SLAM System with Point and Line Features. *Appl. Sci.* **2023**, *15*, 4678.
5. Zhang, Y.; Huang, Y.; Zhang, Z.; Postolache, O.; Mi, C. A vision-based container position measuring system for ARMG. *Meas. Control* **2023**, *56*, 596–605.
6. Duan, H.; Xin, L.; Xu, Y.; Zhao, G.; Chen, S. Eagle-vision-inspired visual measurement algorithm for UAV's autonomous landing. *Int. J. Robot. Autom.* **2020**, *35*, 94–100.
7. Usenko, V.; Demmel, N.; Schubert, D.; Stücker, J.; Cremers, D. Visual inertial mapping with non-linear factor recovery. *IEEE Robot. Autom. Lett.* **2020**, *5*, 422–429.
8. Forster, C.; Pizzoli, M.; Scaramuzza, D. SVO: Fast semi-direct monocular visual odometry. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014.
9. Engel, J.; Koltun, V.; Cremers, D. Direct sparse odometry. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 611–625.
10. Shen, S.; Michael, N.; Kumar, V. Tightly-coupled monocular visual-inertial fusion for autonomous flight of rotorcraft MAVs. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015.
11. Li, M.; Mourikis, A. High-precision, consistent EKF-based visual inertial odometry. *Int. J. Robot. Res.* **2013**, *32*, 690–711.
12. Bloesch, M.; Burri, M.; Omari, S.; Hutter, M.; Siegwart, R. Iterated extended kalman filter based visual-inertial odometry using direct photometric feedback. *Int. J. Robot. Res.* **2017**, *36*, 1053–1072.
13. Qin, T.; Li, P.; Shen, S. VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Trans. Robot.* **2018**, *34*, 1004–1020.
14. Leutenegger, S.; Lynen, S.; Bosse, M.; Siegwart, R.; Furgale, P. Keyframe-based visual-inertial odometry using nonlinear optimization. *Int. J. Robot. Res.* **2014**, *34*, 314–334.
15. Greene, W.; Roy, N. Metrically-scaled monocular slam using learned scale factors. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020.
16. Cui, H.; Tu, D.; Tang, F.; Xu, P.; Liu, H.; Shen, S. Vidsfm: Robust and accurate structure-from-motion for monocular videos. *IEEE Trans. Image Proc.* **2022**, *31*, 2449–2462.
17. Lee, S.; Hwang, S. Elaborate monocular point and line slam with robust initialization. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019.
18. He, Y.; Zhao, J.; Guo, Y.; He, W.; Yuan, K. PI-VIO: Tightly-coupled monocular visual-inertial odometry using point and line features. *Sensors* **2018**, *18*, 1159.
19. Fu, Q.; Wang, J.; Yu, H.; Ali, I.; Guo, F.; He, Y.; Zhang, H. PL-VINS: Real-time monocular visual-inertial SLAM with point and line features. *arXiv* **2020**, arXiv:2009.07462.
20. Lee, J.; Park, S. PLF-VINS: Real-time monocular visual-inertial SLAM with point-line fusion and parallel-line fusion. *IEEE Robot. Autom. Lett.* **2021**, *6*, 7033–7040.
21. Gomez-Ojeda, R.; Briaes, J.; Gonzalez-Jimenez, J. PL-SVO: Semi-direct monocular visual odometry by combining points and line segments. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Republic of Korea, 9–14 October 2016.
22. Lim, H.; Jeon, J.; Myung, H. UV-SLAM: Unconstrained line-based SLAM using vanishing points for structural mapping. *IEEE Robot. Autom. Lett.* **2022**, *7*, 1518–1525.
23. Pumarola, A.; Vakhitov, A.; Agudo, A.; Sanfeliu, A.; Moreno-Noguer, F. PL-SLAM: Real-time monocular visual SLAM with points and lines. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017.
24. Von Gioi, R.; Jakubowicz, J.; Morel, J.; Randall, G. LSD: A fast line segment detector with a false detection control. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 722–732.
25. Mur-Artal, R.; Tardós, J. ORB-SLAM2: An open-source slam system for monocular, stereo, and RGB-D cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262.
26. Akinlar, C.; Topal, C. EDLines: A real-time line segment detector with a false detection control. *Patt. Recog. Lett.* **2011**, *32*, 1633–1642.

27. Liu, Z.; Shi, D.; Li, R.; Qin, W.; Zhang, Y.; Ren, X. PLC-VIO: Visual-Inertial Odometry Based on Point-Line Constraints. *IEEE Trans. Autom. Sci. Eng.* 2021, 19, 1880–1897.
28. Suárez, I.; Buenaposada, J.; Baumela, L. ELSEd: Enhanced line SEgment drawing. *arXiv* 2022, arXiv:2108.03144.
29. Zhao, Z.; Song, T.; Xing, B.; Lei, Y.; Wang, Z. PLI-VINS: Visual-Inertial SLAM based on point-line feature fusion in Indoor Environment. *Sensors* 2022, 22, 5457.
30. Bevilacqua, M.; Berthoumieu, Y. Multiple-feature kernel-based probabilistic clustering for unsupervised band selection. *IEEE Trans. Geosci. Remote Sens.* 2019, 57, 6675–6689.
31. Cipolla, R.; Drummond, T.; Robertson, D. Camera Calibration from Vanishing Points in Image of Architectural Scenes. In *Proceedings of the British Machine Vision Conference 1999, Nottingham, UK, 13–16 September 1999; Volume 38*, pp. 382–391.
32. Chuang, J.; Ho, C.; Umam, A.; Chen, H.; Hwang, J.; Chen, T. Geometry-based camera calibration using closed-form solution of principal line. *IEEE Trans. Image Process.* 2021, 30, 2599–2610.
33. Liu, J.; Meng, Z. Visual SLAM with drift-free rotation estimation in manhattan world. *IEEE Robot. Autom. Lett.* 2020, 5, 6512–6519.
34. Camposeco, F.; Pollefeys, M. Using vanishing points to improve visual-inertial odometry. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Seattle, WA, USA, 26–30 May 2015.
35. Li, Y.; Yunus, R.; Brasch, N.; Navab, N.; Tombari, F. RGB-D SLAM with structural regularities. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China, 30 May–5 June 2021.
36. Li, X.; Zhu, L.; Yu, Z.; Guo, B.; Wan, Y. Vanishing Point Detection and Rail Segmentation Based on Deep Multi-Task Learning. *IEEE Access* 2020, 8, 163015–163025.
37. Kim, P.; Coltin, B.; Kim, H. Low-drift visual odometry in structured environments by decoupling rotational and translational motion. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, Australia, 21–25 May 2018.
38. Li, Y.; Brasch, N.; Wang, Y.; Navab, N.; Tombari, F. Structure-SLAM: Low-drift monocular SLAM in indoor environments. *IEEE Robot. Autom. Lett.* 2020, 5, 6583–6590.
39. Zou, D.; Wu, Y.; Pei, L.; Ling, H.; Yu, W. StructVIO: Visual-inertial odometry with structural regularity of man-made environments. *IEEE Trans. Robot.* 2019, 35, 999–1013.
40. Lu, X.; Yao, J.; Li, H.; Liu, Y.; Zhang, X. 2-line exhaustive searching for real-time vanishing point estimation in manhattan world. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, Santa Rosa, CA, USA, 24–31 March 2017.
41. Peng, X.; Liu, Z.; Wang, Q.; Kim, Y.; Lee, H. Accurate visual-inertial slam by manhattan frame re-identification. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Prague, Czech Republic, 27 September–1 October 2021.