

Research Article

Mobile Robot Simultaneous Localization and Mapping Based on a Monocular Camera

Songmin Jia,^{1,2} Ke Wang,^{1,2} and Xiuzhi Li^{1,2}

¹College of Electronic Information & Control Engineering, Beijing University of Technology, Beijing 100124, China

²Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing 100124, China

Correspondence should be addressed to Songmin Jia; jsm@bjut.edu.cn

Received 25 December 2015; Revised 11 April 2016; Accepted 8 May 2016

Academic Editor: Shahram Payandeh

Copyright © 2016 Songmin Jia et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a novel monocular vision-based SLAM (Simultaneous Localization and Mapping) algorithm for mobile robot. In this proposed method, the tracking and mapping procedures are split into two separate tasks and performed in parallel threads. In the tracking thread, a ground feature-based pose estimation method is employed to initialize the algorithm for the constraint moving of the mobile robot. And an initial map is built by triangulating the matched features for further tracking procedure. In the mapping thread, an epipolar searching procedure is utilized for finding the matching features. A homography-based outlier rejection method is adopted for rejecting the mismatched features. The indoor experimental results demonstrate that the proposed algorithm has a great performance on map building and verify the feasibility and effectiveness of the proposed algorithm.

1. Introduction

To successfully accomplish autonomous or semiautonomous tasks to assist daily human activities, building the 3D geometry map of the environment is becoming one of the fundamental issues for mobile robotics. In computer vision community, tracking and mapping of the sparse tracked features can be easily conducted in real time by using a hand-held camera. Existing Structure-from-Motion (SfM) [1] and filter-based Simultaneous Localization and Mapping (SLAM) [2] approaches have been investigated in recent years in order to concurrently estimate the pose of a freely moving camera and the sparse spatial structure of the environment [3].

Computer vision-based approaches offer substantial advantages for robot localization and environment recognition. Therefore, a lot of researches were devoted to extend the vision technique for mobile robot platform. Recently, it has been shown that it is possible to estimate the trajectory using a monocular camera in real time. The typical filter-based vision SLAM is proposed by Davison with a single camera. In Davison's work, a probabilistic filtering was used for features of tracking and mapping procedures in a consistent scene map and the camera pose was estimated with

a relative coordination in real time [4]. Royer et al. proposed a monocular vision system for mobile robot localization and autonomous navigation with the prevailing SfM SLAM [5]. In Royer et al.'s work, a sparse 3D map was built with a video sequence for the learning step. Then the sparse map was used to localize the robot for its learning path or a slightly different path. In particular, Klein and Murray proposed a famous SfM-based vision SLAM which is named PTAM (parallel tracking and mapping) [6, 7]. In the PTAM, the mapping and tracking were split into two separate tasks and performed in parallel threads. In the tracking thread, the robust pose estimation of camera was calculated with the feature detection and matching. In the mapping thread, the features which are achieved by epipolar searching procedure were triangulated for map building. Recently, the PTAM algorithm has been widely applied for initial guess for some dense mapping systems [8, 9] or the localization method for robotic systems [10–12]. Forster et al. proposed a semidirect monocular visual odometry algorithm which is somehow different from PTAM [13]. In their work, the feature extraction was only required when a key frame is selected to initialize new 3D points. In the tracking step, the correspondence of feature was an implicit result of direct motion estimation as in [14]. For the outlier

measurement of triangulated features' depth, a Bayesian filter was used to estimate the depth at feature locations.

Spirited by above-mentioned works, this paper aims to propose a robust algorithm for a mobile robot based on an improved PTAM. The main contributions of this paper are that an initialization method is proposed by using ground features and a homography-based constraint is used for mismatched features rejection. Assume that a camera is fixed on a mobile robot which is moving on a plane at the beginning. In the initialization procedure, the mobile robot pose is roughly localized by using a homography-based pose estimation method with features tracked by frames. When the distance of mobile robot is about up to 5 cm, the ground features-based pose estimation is performed to achieve the accurate localization of robot. And the pose of camera is calculated by using the translation between camera and robot. Then, the initial map is established by triangulating those matched features. After this initialization procedure, the algorithm begins to track the robot by using a weighted projection error function. When a key frame is achieved, the mapping procedure is performed to achieve the matched features by using epipolar searching and all the matched features will be tested by a homography-based constraint for outlier rejection. To optimize the generated map, local and global bundle adjustments are also adapted as in PTAM algorithm. The indoor experimental results demonstrate the feasibility and effectiveness of the proposed algorithm.

The rest of the paper is organized as follows: in Section 2, each part of the proposed algorithm is explained in detail. Section 3 presents the experimental results of our system in real environments. Finally, conclusions are drawn and future work is outlined in Section 4.

2. Algorithm Architecture

The overall structure of the proposed algorithm is illustrated in Figure 1. Following the structure of the original PTAM algorithm, the proposed algorithm is divided into two independent threads which are the tracking thread and the mapping thread. In the tracking thread, a pre-calibrated camera which is fixed on the mobile robot is capturing the images from the experiment scene. Then, a ground feature-based pose estimation algorithm is employed to achieve a high quality initial map for further camera tracking procedure. After the initialization, the robot pose is estimated by minimizing a weighted projection error-based energy function. When a key frame is selected, the mapping procedure will start in the other thread. The new map points are detected by using the epipolar searching between the new coming key frame and its nearest key frame. To get rid of the mismatched features, a homography-based constraint is utilized to reject the outliers and achieve a clear and high quality 3D sparse map building. Finally, the local and global bundle adjustments are used to achieve an optimized map.

2.1. Parameter Definition. Let $p = (x, y, z)^T$ represent the coordinate of a 3D point defined in the camera reference frame and $u_n = (x_n, y_n)^T$ denotes the normalized image

projection coordinate. For a pinhole camera, projection of a 3D point p onto the image plane appears as

$$u_n = K\pi(p), \quad (1)$$

where $\pi(p) = (x/z, y/z, 1)$ is the projection function. The term K specifies the intrinsic parameter matrix; that is,

$$K = \begin{bmatrix} f_x & 0 & u_x \\ 0 & f_y & u_y \\ 0 & 0 & 1 \end{bmatrix}, \quad (2)$$

where (f_x, f_y) is the focal length expressed in units of horizontal and vertical pixels, whereas (u_x, u_y) denotes the principal point coordinates of the image plane. It is worth noting that the above linear model fails to describe a wide-angle camera. In this work, conforming to the mathematical model proposed by Devernay and Faugeras [15], geometric mapping constraints between the distorted image projection (with subscript d) and the corresponding undistorted location (with subscript u) are formulated as

$$u_d = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \begin{bmatrix} f_u & 0 \\ 0 & f_v \end{bmatrix} \cdot \frac{r_d}{r_u} \cdot u, \quad (3)$$

$$r_d = \frac{1}{\omega} \arctan\left(2r_u \tan \frac{\omega}{2}\right).$$

r is the radial distance; that is, $r_u = \|u\|$. We use the pose of camera with respect to the world frame of reference as

$$T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}, \quad (4)$$

where $T \in SE(3)$ is a 4×4 matrix describing the camera pose for each frame relative to the world, R denotes the rotation matrix, and t stands for the translation vector. T transforms the point from the camera's frame of reference to that of the world, such that $p_c = Tp_w$.

Assume that the camera is fixed on the mobile robot with a constant transformation described by transformation T_{cr} as is shown in Figure 1. Let T_{rs} be the transformation describing the relative motion of the robot within a time step when its camera captures two images. Since T_{cr} is a constant, the relative motion of the camera between the captured two images in its own frame of reference T_{cs} is related to the motion of the robot and can be parameterized by planar motion of the robot as follows:

$$T_{rs} = T_{rc}T_{cs}T_{cr}, \quad (5)$$

where $T_{cr} = T_{rc}^{-1}$.

2.2. Initialization of Algorithm. As the ground features are sometimes much richer than the upper space in the indoor environment, a ground feature-based localization algorithm is adopted for initializing the improved PTAM algorithm. Two captured images (live I_c and reference I_r) of the same ground plane are related by a plane homography matrix H .

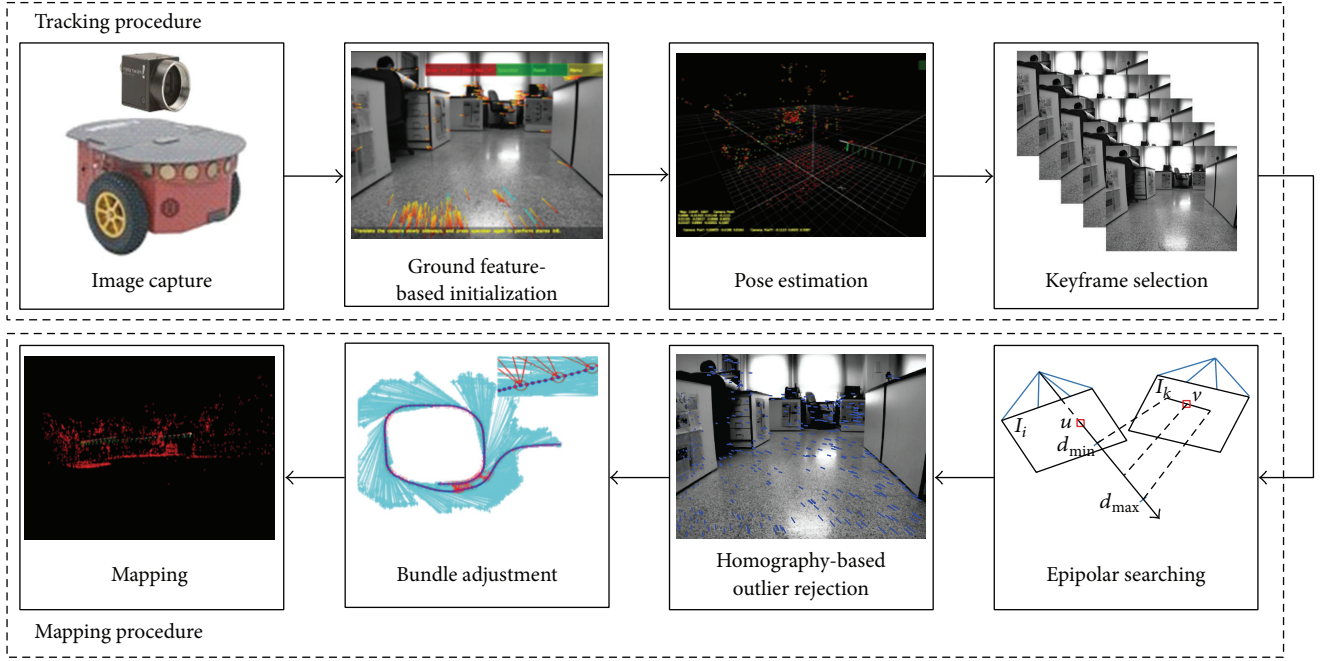


FIGURE 1: The illustration of the proposed algorithm.

For two corresponding features u and v , the homography matrix transforms feature coordinates in the reference image into the pixels in the live image and depends on the camera motion T_{cs} and the parameters n_d of the ground plane in the reference camera frame:

$$u = \alpha H v, \quad (6)$$

$$H = K T_{cr} T_{rs}(\mathbf{x}) T_{rc} (I | n_d)^T K^{-1},$$

where α is the scale factory. n_d and T_{rc} are easily precalculated by using the calibration method for camera.

The pose of camera can be achieved by minimizing the following energy function:

$$\arg \min_{\mathbf{x}} \sum \|r_p(\mathbf{x})\|, \quad (7)$$

where $r_p = \pi(H(\mathbf{x})u) - v$ is the homography-based projection error.

As the nonground features have great influence on the pose estimation, a ground feature detection and pose estimation algorithm is expressed as (8) according to the homography-based pose estimation algorithm and a 3D scene flow depth estimation algorithm:

$$\arg \min_{\mathbf{x}, d} \sum L(d(u) - d_g(u)) \|r_p(\mathbf{x})\| + \|r_d(u, d)\|, \quad (8)$$

$$L(x) = \begin{cases} 1 & |x| \leq \sigma \\ 0 & \text{others,} \end{cases}$$

where $d_g = 1/n^T u$ is the depth of the ground feature; d is the depth estimated by a 3D scene flow depth estimation algorithm. $L(x)$ indicates whether the feature is placed on

the ground. Only when $|d_i - d_g| \leq \delta$, $L(x) = 1$. $r_d = K\pi(T_{cs}(\mathbf{x})p(u, d)) - v$ measures the projection error driven by the depth of features. $p(u, d)$ is the coordinate of the 3D point corresponding to the feature u with the depth d .

To optimize this proposed ground feature detection model, an alternating scheme is utilized to update either \mathbf{x} or d in every iteration. The full procedure in iteration is shown as follows:

(a) Fix d ; solve

$$\arg \min_{\mathbf{x}} \sum L(d - d_g) \|r_p(\mathbf{x})\|. \quad (9)$$

To optimize this energy function, the first-order Taylor expansion of $r_p(\mathbf{x})$ is performed around $\mathbf{x} = \mathbf{x}_0$. Then, the error function becomes

$$r_p(\xi) = r_p(\mathbf{x}_0) + J\xi, \quad (10)$$

where $J_p = (\partial r_p / \partial \mathbf{x})|_{\mathbf{x}_0}$.

Finally, a Gauss-Newton step is obtained to efficiently optimize that error function iteratively. The solution of energy function becomes

$$\sum J_p^T J_p \xi = -\sum J_p^T r_p(\mathbf{x}_0), \quad (11)$$

$$\xi = -(\sum J_p^T J_p)^{-1} \sum J_p^T r_p(\mathbf{x}_0).$$

The updating procedure of robot pose is applied as follows:

$$T = T_{\text{inc}}(\mathbf{x})T, \quad (12)$$

where T is the last estimate of the solution and $T_{\text{inc}}(\mathbf{x})$ is a small update to the estimate parameterized by $\mathbf{x} \in \text{se}(2)$.

In this approach, we are required to evaluate the partial derivatives of the cost function at the solution $\mathbf{x} = \mathbf{x}_0$. The derivation of the error function is calculated in each feature point by partial derivatives:

$$\begin{aligned} \frac{\partial r_p(\mathbf{x})}{\partial x_i} &= \frac{\partial \pi(p)}{\partial p} \frac{\partial H(\mathbf{x})}{\partial x_i} p, \\ \frac{\partial H(\mathbf{x})}{\partial x_i} &= K T_{rc} T_{rs} \frac{\partial T(\mathbf{x})}{\partial x_i} T_{rc} (I | n_d)^T K^{-1}, \end{aligned} \quad (13)$$

where x_i is the i th generator of the $se(2)$ group.

(b) Fix \mathbf{x} ; solve

$$\arg \min_d \sum \|r_d(u, d)\|. \quad (14)$$

The energy function can be solved by iterating the following equations:

$$\begin{aligned} \Delta u_i &= J_d \nabla d, \\ d &= d + \nabla d, \end{aligned} \quad (15)$$

where J_d is the Jacobian for the projection function evaluated at point u ; $J_d = (\partial \pi(p)/\partial p)(\partial p/\partial d)$. Generally, it only needs 3~5 iterations to achieve the convergent solution.

(c) Revalue δ by using $\delta = \max(\gamma, \beta\delta)$.

(d) The iteration will not stop until the ground features are no longer changed or the maximum iteration is arrived at.

(e) When all the ground features are detected, all the ground features are used to achieve more accurate pose estimation with an m-estimation scheme of (9). The weighted energy function is described as follows:

$$\min_{\mathbf{x}} \sum w(r_p) \|r_p(\mathbf{x})\|, \quad (16)$$

where w is Tukey's biweight function [16] for r_p and

$$\begin{aligned} x &= \frac{|x_i|}{c x_m}, \\ w(x) &= \begin{cases} \left(1 - \left(\frac{x}{b}\right)^2\right)^2 & |x| \leq b \\ 0 & \text{others,} \end{cases} \end{aligned} \quad (17)$$

where x_m is the median value of x , $c = 1.4826$ is the robust standard deviation, and $b = 6.6851$ is Tukey specific constant.

Figure 3 illustrates the initialization of the proposed algorithm. Firstly, the algorithm is tracking the features using a patch-based method as is shown in Figure 3(a). All tracked features are regarded as the ground features for raw pose estimation by solving (9). When the translation of mobile robot is about up to 5 cm, the ground feature detection and pose

estimation algorithm is performed for the accurate localization of mobile robot. In Figure 3(b), all the ground features are detected by using the proposed algorithm, and reliable and accurate pose estimation is achieved by using the m-estimation scheme. With this accurate localization of mobile robot, all matched features are triangulated for map building and an initial map is established after the epipolar searching procedure, as shown in Figures 3(c) and 3(d). It demonstrates that the proposed algorithm is able to detect all the ground features and achieve high accurate pose estimation during the initialization.

2.3. Tracking Procedure. Accurate camera pose determination is necessitated in a consistent monocular map building, and the goal of this process is to find a rotation matrix R and translation vector t which transform the current camera coordinate frame to that of the global world. In this work, the optimal pose is derived by iterating a robust objective function which minimizes the reprojection error on the image plane used in PTAM algorithm:

$$\xi = \arg \min_{\xi} \sum_i w(e_i) \|e_i\|, \quad (18)$$

where e_i is the projection error. $w(e_i)$ stands for Tukey's biweight function.

After intense experimental study, we come to realize that the quality of map initialization has a significant impact on the subsequent mapping. Prompted by this requirement, we initiate the 3D point cloud before the dynamic tracking. The method in Section 2.2 is performed for initialization. During this initialization procedure, an image pair is thus gathered at different camera locations which are corresponding to the robot moving and estimated by employing a homography-based method. When the distance of robot moving is about 5 cm and an accurate localization of robot is achieved during the initialization, a good quality initial map guess is provided by triangulating the correspondent points for further tracking and sparse mapping procedure.

Once tracking starts, we attempt to track the FAST-10 corners [17] within the impeding images and they are employed as visual features for tracking in data association procedure. Spatial 6D trajectory of the moving camera could be derived by SFM. As correct interframe correspondence is of vital importance, FAST-10 features are extracted and matched on a four-level coarse-to-fine architecture to boost tracking performance. An 8×8 -pixel patch template is employed for matching the corner features and the most appealing matched pair is selected by searching within the circular candidate region with the smallest difference ZMSSD (zero-mean SSD) [18] scores. The features on coarse levels of the image pyramid are searched to roughly estimate the relative pose with (18), and then a final accurate pose is optimized by using a thousand of feature correspondences. During the experiment, the tracking is executed with a key frame-based strategy for subsequent incremental mapping. A key frame is selected by taking into account such factors as tracking quality, number of skip frames, proportion of new features, and distance away from the nearest key point.

TABLE 1: The deviation evaluation of original PTAM and proposed algorithm to ground only PTAM.

		x	y	z	Pitch	Yaw	Roll
Proposed algorithm	Mean	-0.0055	0.0052	0.0015	-0.0003	-0.0002	-0.0003
	Variance	0.0025	0.0048	0.0015	0.0006	0.0009	0.0004
Original PTAM	Mean	-0.0231	0.1835	0.0104	-0.0035	0.0057	0.0044
	Variance	0.1166	0.2207	0.1540	0.0272	0.0159	0.0115

2.4. Mapping Procedure. Based on the pose estimation by image features, the matched feature points are triangulated into 3D point cloud. After the sparse map initialization stated previously, the sparse mapping thread begins to query for incorporation of new key frames. Once a new key frame is successfully selected, the extracted FAST features are evaluated by using Shi-Tomas score to find out the most salient points in the new key frame. As Euclidean transformation between current key frame and its closest neighbor, which is reported in tracking routine, is available, new correspondent features can be found in adjacent key frame by using epipolar constraint and ZMSSD criterion is applied for those remaining unmatched feature points in current key frame. Then those new map points are incorporated into the map after triangulation of new feature pairs.

It is worth noting that projection of a 3D point in the sparse map is generally inconsistent with the corresponding feature point due to inaccuracy involved in feature finding, pose estimation, and so forth. A small region around the corresponding features can usually be found in epipolar searching procedure between key frames. With this association among the key frames, a refine map is optimized by using a bundle adjustment routine [19] to simultaneously optimize the pose for all key frames and all map points. In practice, this procedure becomes increasingly expensive in computation as map size and the number of key frames increase. For this reason, a local bundle adjustment is also adopted to perform in this thread. The most recent key frame and its closest neighbors and all of the map points in their view are adjusted by using the local bundle adjustment. According to this local and global bundle adjustment mechanism, the key frame can be integrated in a very short possible time, and it also allows a reasonable rate of camera tracking.

2.5. Homography-Based Outlier Rejection. In the mapping thread, an epipolar searching procedure is employed to achieve some new matched features. The matched features are searched by finding a minimum difference of ZMSSD among features which are around the epipolar. Ideally, the ZMSSD value between two matched patches should be very small. In particular, the ZMSSD which is equal to 0 means that two matched patches are perfectly fitted. As ZMSSD is not an invariant with the rotations, some mismatched features are triangulated and added into map in PTAM. Although a few mismatched features will be rejected during the bundler adjustment optimization, most of them are hardly to be removed. As a projection error-based pose estimation algorithm is utilized in PTAM, the noise will have some effect on the accuracy of localization.

To analyze the performance of features matching in PTAM, the features are detected in the three typical images which are shown in Figure 4(a) by using the FAST-10 algorithm. The ZMSSDs between the original image patch and warped patch which, respectively, rotates and translates along x -axis, y -axis, and z -axis are calculated and drawn in Figures 4(b)–4(d) for all features. The ZMSSD values of some features with different rotation or translation are easily confused for feature matching during the patch searching. Moreover, the epipolar searching procedure is generally hard to establish one-one mapping for feature matching with such confused ZMSSD evaluation. Therefore, a noised map and the low accurate pose estimation are sometimes achieved by the original PTAM algorithm.

To efficiently establish a correct feature matching relationship, a homography-based constraint is adopted for those new detected features. After epipolar searching procedure in i th image level, the homography H_i is estimated by using a RANSAC approach with all the detected features. Only when the projection error r_j is smaller than the threshold δ_i is the detected feature u_j accepted as good matched feature. Consider

$$s(u_j^i) = \begin{cases} r_j \leq \sigma_i & \text{accept} \\ r_j > \sigma_i & \text{reject} \end{cases} \quad (19)$$

$$r_j = \pi(H_i v_j^i) - u_j^i,$$

where v_j^i is the reference feature of u_j^i . Generally, the homography in the higher level of the image pyramid can be approximated by the lower one with the following equation:

$$v_j^{i-1} = \begin{bmatrix} \frac{1}{s} & & \\ & \frac{1}{s} & \\ & & 1 \end{bmatrix} v_j^i,$$

$$H_i \begin{bmatrix} \frac{1}{s} & & \\ & \frac{1}{s} & \\ & & 1 \end{bmatrix} v_j^i = s \begin{bmatrix} \frac{1}{s} & & \\ & \frac{1}{s} & \\ & & 1 \end{bmatrix} u_j^i, \quad (20)$$

$$H_i = \begin{bmatrix} \frac{1}{s} & & \\ & \frac{1}{s} & \\ & & 1 \end{bmatrix} H_{i-1} \begin{bmatrix} s & & \\ & s & \\ & & 1 \end{bmatrix},$$

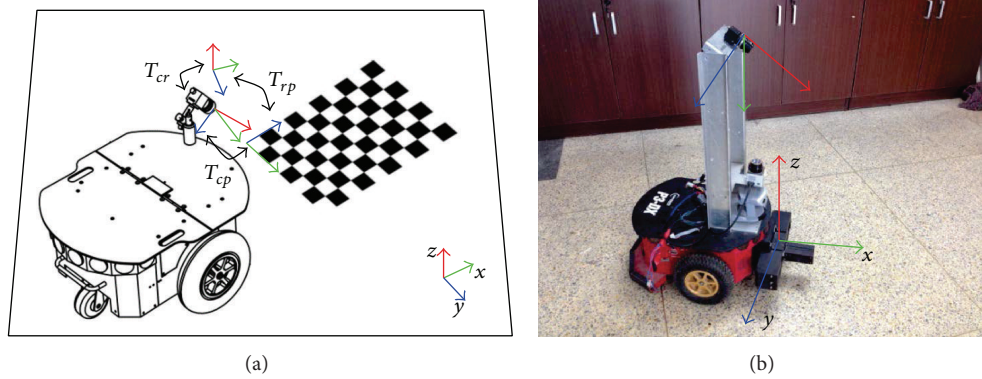


FIGURE 2: The parameter calibration for mobile robot: (a) is the procedure of calibration of plane parameter, and (b) is the robot platform.

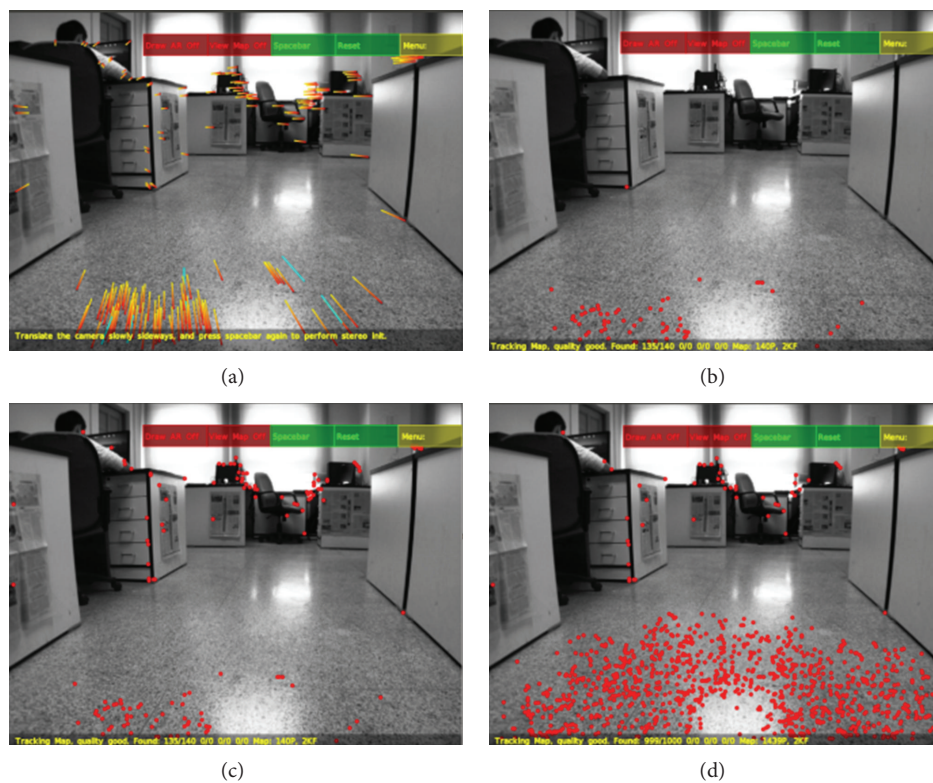


FIGURE 3: The procedure of initialization: (a) illustrates the tracking of features, the detected ground features are shown in (b), (c) shows all the matched features, and (d) demonstrates the initial map after epipolar searching.

where s is the down-sample rate of image pyramid. During the experiment, the threshold δ_i in the high level of image pyramid is also decreased by the scale of image down-sample rate.

Figure 5 is a compare result between the original PTAM algorithm and the improved PTAM algorithm. The experimental environment is shown in Figure 5(a). Figure 5(b) shows the map generated by the original PTAM algorithm. 7,680 feature points are achieved in the established map and 6 key frames are generated during the camera moving. There are some mismatched features which have been triangulated into the map. Though a few of the mismatched features are

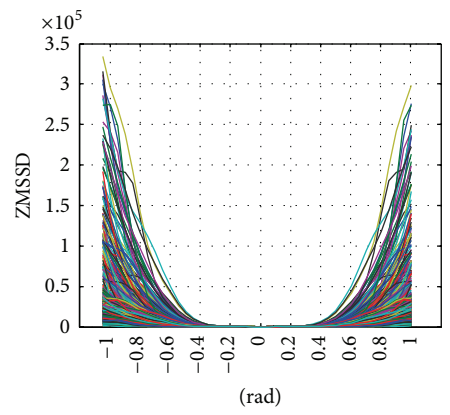
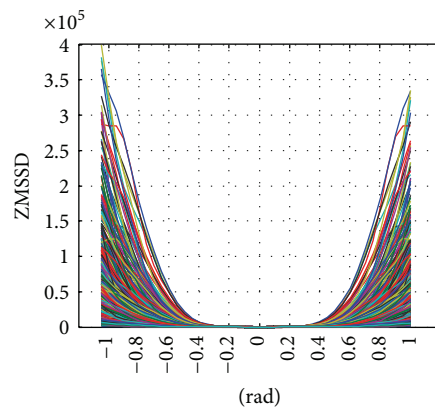
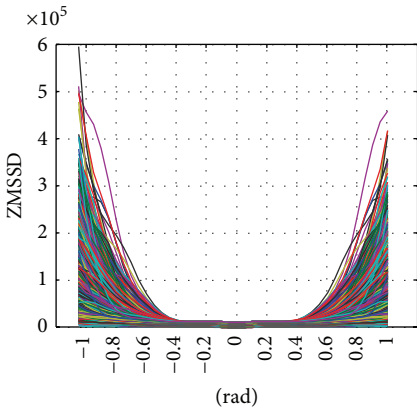
rejected after bundle adjustment optimization, some mismatched features still exist which would have an effect on the localization of PTAM algorithm. Figure 5(c) demonstrates the map established by using PTAM with the homography-based constraint. More than 1500 feature points are accepted and added to the map. And the mismatched features are rejected by using the homography-based constraint.

3. Experimental Result

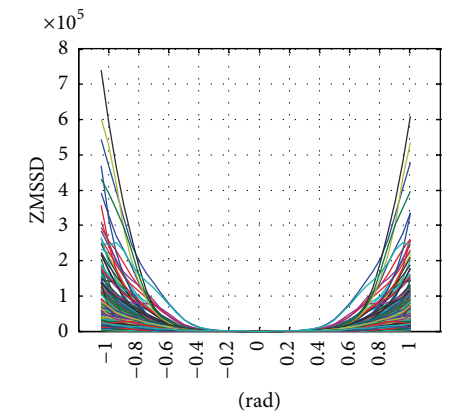
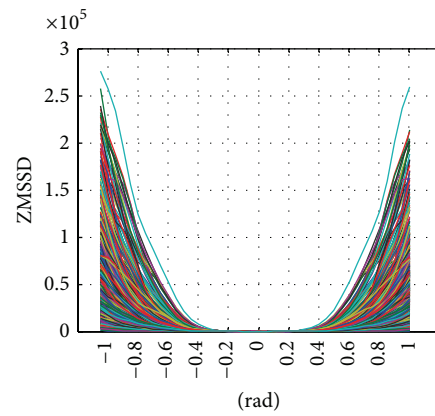
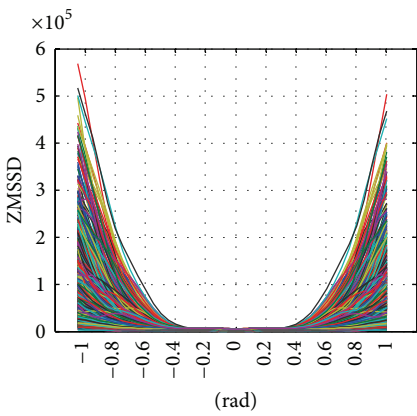
Indoor experiments are conducted to verify the validity and accuracy of the algorithm. Our hardware system setup



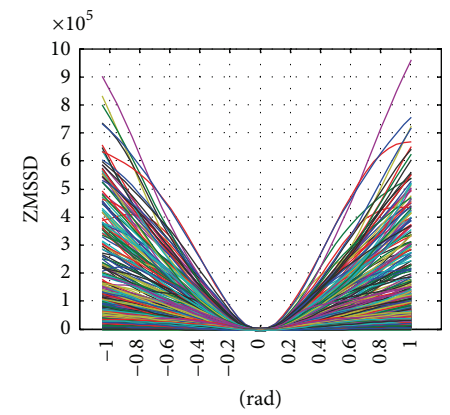
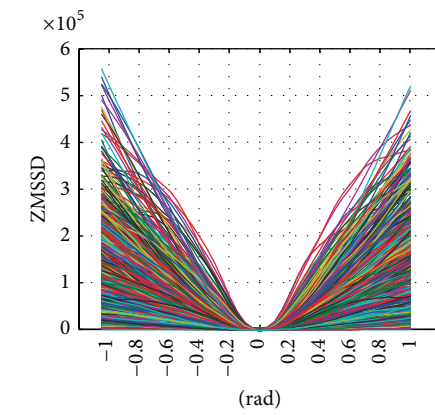
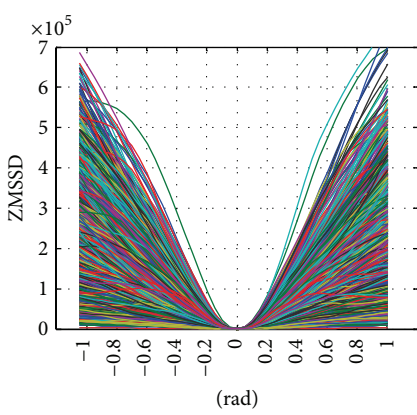
(a)



(b)



(c)



(d)

FIGURE 4: Continued.

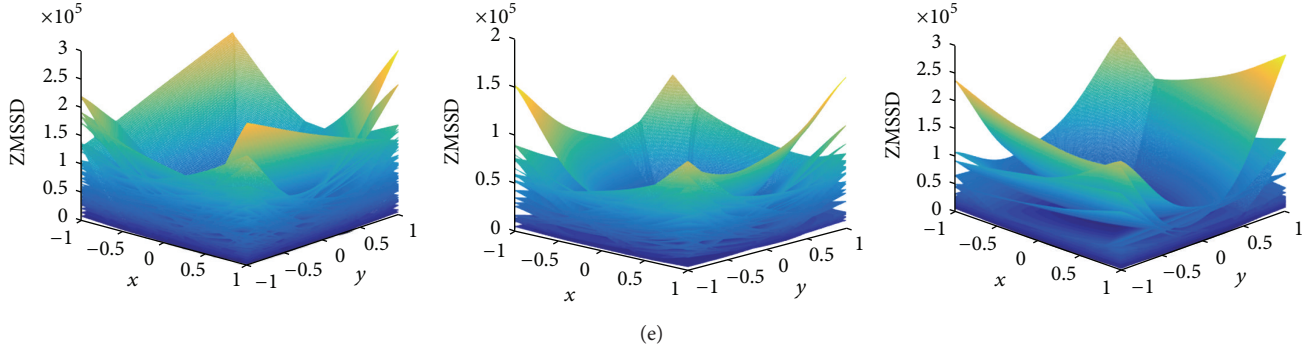


FIGURE 4: The ZMSSDs for rotation and translation of features: (a) shows the images for feature detection; (b), (c), and (d) are the ZMSSDs with rotations around x -axis, y -axis, and z -axis; (d) illustrates the ZMSSDs with the translations along x -axis and y -axis.

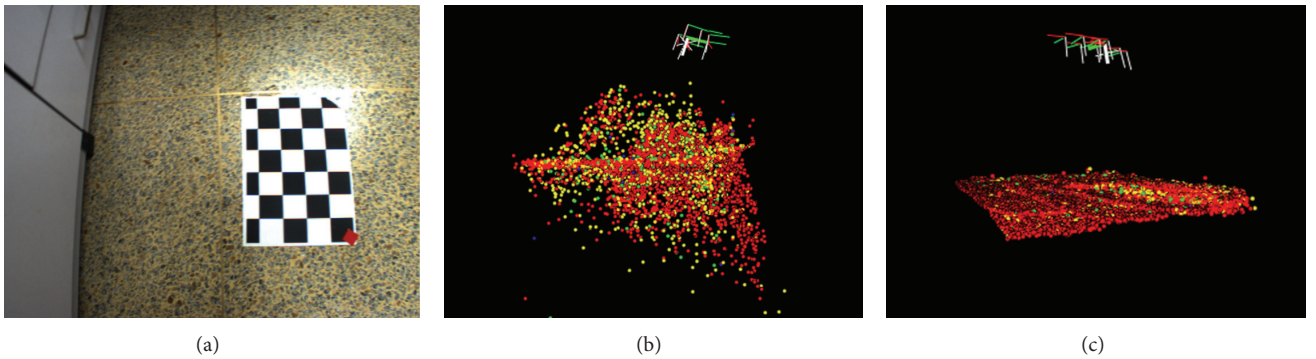


FIGURE 5: A compare result of the original PTAM and the improved PTAM algorithm: (a) is the experiment environment, (b) shows the map generated by the original PTAM, and (c) shows the map generated by the improved PTAM.

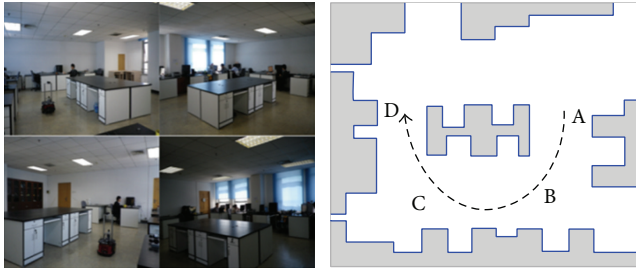


FIGURE 6: The experimental environment.

consists of a Pioneer3-DX platform and a monocular vision system which is depicted in Figure 2. The robot platform is qualified for the indoor research activities like mapping, obstacle avoidance, path planning, and so forth. A forward looking Point Grey Flea2 camera (shown in Figure 1) is fixed on the mobile robot. The focal length of the camera is about 3.5 mm and the field of view is about 65° . The camera is connected via 1394 B to a 2.8 GHz Intel i5 computer with 4 G RAM. The experiments are performed in an indoor environment as shown in Figure 6.

During the experiment, the mobile robot is running following the path from A to D, while the camera is capturing the images. Firstly, the algorithm is initialized by using

the proposed ground features-based method. During the initialization procedure, the features are tracked frame by frame. All the features are regarded as the ground features and the coarse localization of mobile robot is estimated by (9). When the distance of robot moving is about up to 5 cm, the ground features detection procedure is performed to achieve accurate localization for the mobile robot. With the translation between the mobile robot and camera, the estimated pose of camera is calculated for initializing the mapping and epipolar searching. Then, all matched features are triangulated and added to the map for further tracking procedure. After the initialization, the algorithm begins to estimate its pose by minimizing a weighted projection error function and the map is updated with the epipolar searching after a key frame is achieved. To evaluate the proposed method, the PTAM algorithm which only uses the ground features in tracking procedure is considered as the ground truth during the experiment. In the ground only PTAM algorithm, the detected features are triangulated to achieve the 3D coordination p_i . And the features are accepted in the algorithm only when $0.95 \leq \|n_d^T p_i\| \leq 1.05$. As the initialization of the original PTAM algorithm is sometimes unsuitable during the experiment, the ground feature-based pose estimation is employed to initialize the algorithm.

Figure 7 is the mapping result of the proposed algorithm and the original PTAM algorithm. Figures 7(a) and 7(b) are

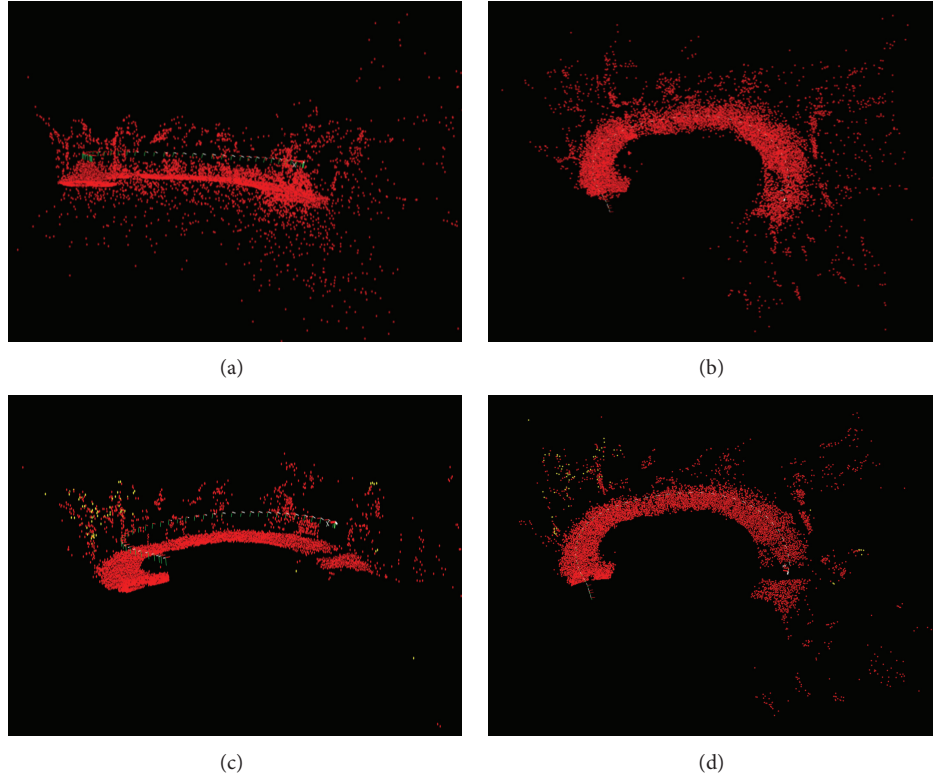


FIGURE 7: The mapping result: (a) and (c) are generated by original PTAM algorithm and proposed algorithm; (b) and (d) are the downlooking of (a) and (c).

the map generated by the original PTAM algorithm in the different viewpoints. Some mismatched features have been added to the map by using the epipolar searching procedure. It would have much influence on the pose estimation with mismatched features. Even worse is the fact that the tracking procedure would make some ridiculous mistakes and fail to track the features, which is hardly to be recovered. To deal with these mismatched features, a homography-based constraint is employed in our improved algorithm. After the epipolar searching procedure, all the matched features which are detected between the new coming key frame and its nearest key frame are used to estimate homography with a RANSAC scheme. The features are filtered with the distance of homography projection errors. In Figures 7(c) and 7(d), the mismatched ground features are rejected and a clearness map is generated by using the proposed method. The experiment demonstrates that the proposed algorithm is able to achieve a clear and high quality 3D feature map.

Figure 8 shows the compare result of the localization among the original PTAM, ground only PTAM, and our proposed algorithm during the mapping experiment. As is shown in Figure 8, the estimated parameters x , y , z , $pitch$, yaw , and $roll$ of the proposed algorithm and ground only PTAM are highly consistent with each other. But in the original PTAM algorithm, the estimated parameters deviate far from the ground only PTAM algorithm due to the noised map. Table 1 illustrates the mean and the variance of the error which is achieved by the original PTAM and the proposed algorithm. The estimated trajectories of three methods are

shown in Figure 9. The proposed algorithm is highly consistent with the ground only PTAM. All above experimental results demonstrate that the proposed algorithm has a high performance on mismatching features rejection and pose estimation. Those results also verify the feasibility and effectiveness of the proposed algorithm.

4. Conclusions

This paper proposes a parallel tracking and mapping algorithm with a monocular camera for mobile robot. The camera is fixed on the mobile robot with a consistent translation, and the robot is moving in an indoor experiment. A ground feature-based pose estimation method is employed to initialize the algorithm. And an initial map is built by triangulating the matched features for further tracking procedure. To establish a sparse 3D map, an epipolar searching procedure is utilized for finding the matching features when it comes to a key frame. Moreover, a mismatched feature rejection method is adopted based on homography for achieving a clearness map. The indoor experimental results demonstrate that the proposed algorithm has a high performance on sparse feature-based mapping and verify the feasibility and effectiveness of the proposed algorithm.

As future work, we will further improve the performance of the proposed algorithm to deal with the mapping in the lighting changed scene. And we also plan to find way to extend our work to a dense visual SLAM system with a monocular camera.

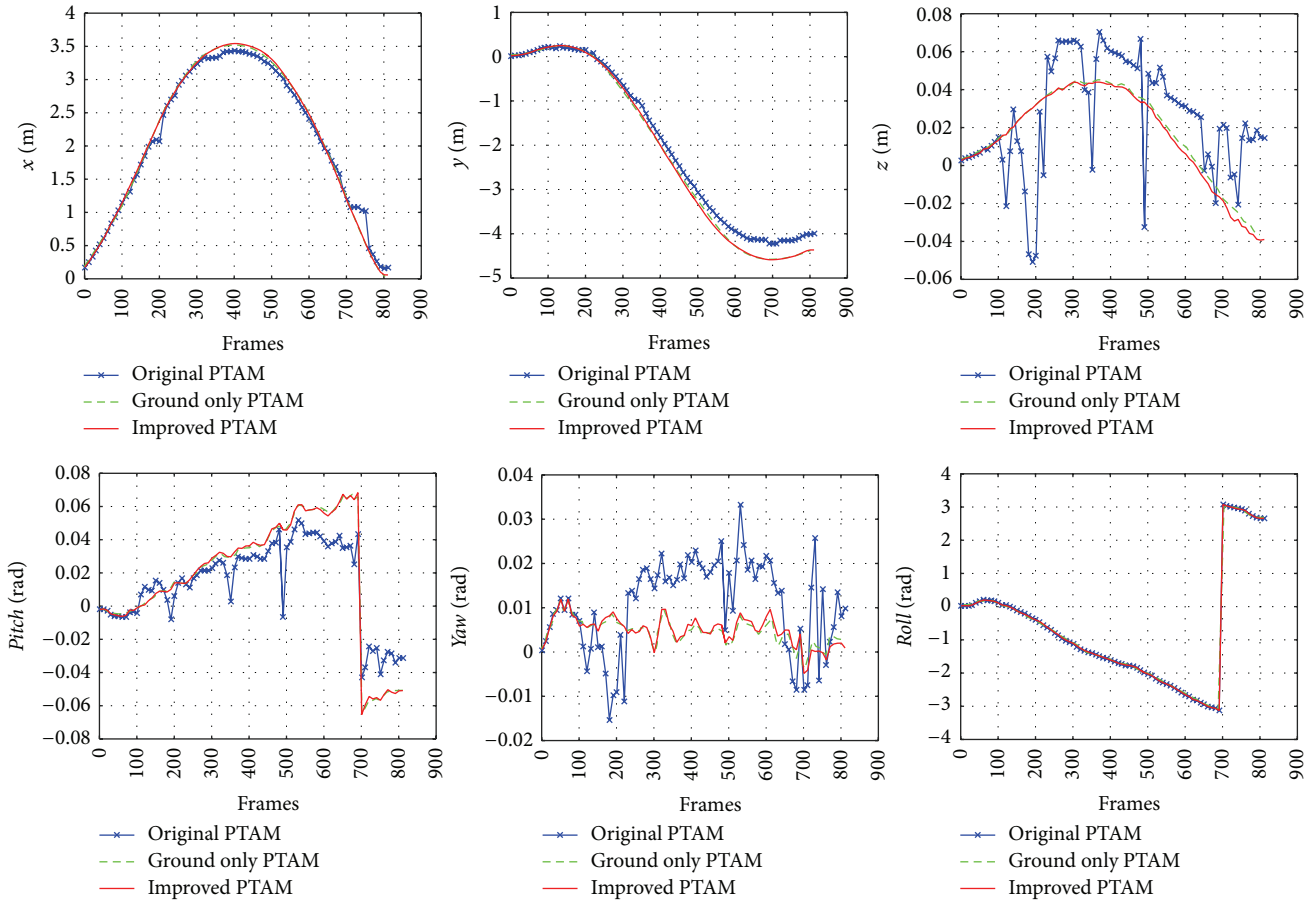


FIGURE 8: The compare result among the original PTAM, ground only PTAM, and our proposed algorithm.

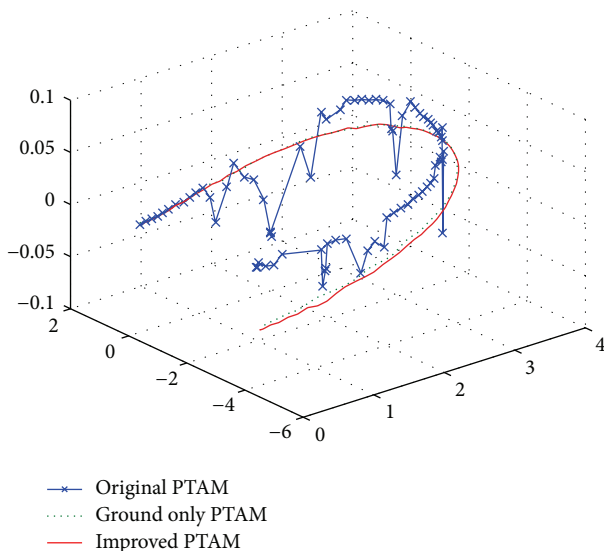


FIGURE 9: The estimated trajectories of mobile robot.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

The research work is financially supported by Promotion Project on Intelligent Robot in Beijing University of Technology “Key Technologies on Modularization and Practicalization of Intelligent Service Robot” and the Natural Science Foundation of China (61175087 and 61105033).

References

- [1] M. Tomono, “3-D localization and mapping using a single camera based on structure-from-motion with automatic baseline selection,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '05)*, pp. 3342–3347, April 2005.
- [2] A. J. Davison, *Mobile Robot Navigation Using Active Vision*, University of Oxford, Oxford, UK, 1998.
- [3] H. Strasdat, J. M. M. Montiel, and A. J. Davison, “Visual SLAM: why filter?” *Image and Vision Computing*, vol. 30, no. 2, pp. 65–77, 2012.
- [4] A. J. Davison, “Real-time simultaneous localisation and mapping with a single camera,” in *Proceedings of the 9th IEEE International Conference On Computer Vision*, pp. 1403–1410, October 2003.
- [5] E. Royer, M. Lhuillier, M. Dhome, and J.-M. Lavest, “Monocular vision for mobile robot localization and autonomous

- navigation,” *International Journal of Computer Vision*, vol. 74, no. 3, pp. 237–260, 2007.
- [6] G. Klein and D. Murray, “Parallel tracking and mapping for small AR workspaces,” in *Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '07)*, pp. 225–234, Nara, Japan, November 2007.
- [7] G. Klein and D. Murray, “Improving the agility of keyframe-based SLAM,” in *Proceedings of the 10th European Conference on Computer Vision (ECCV '08)*, pp. 802–815, October 2008.
- [8] J. Stühmer, S. Gumhold, and D. Cremers, “Parallel generalized thresholding scheme for live dense geometry from a handheld camera,” in *Trends and Topics in Computer Vision*, pp. 450–462, Springer, Berlin, Germany, 2012.
- [9] R. A. Newcombe and A. J. Davison, “Live dense reconstruction with a single moving camera,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 1498–1505, June 2010.
- [10] M. Blösch, S. Weiss, D. Scaramuzza, and R. Siegwart, “Vision based MAV navigation in unknown and unstructured environments,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '10)*, pp. 21–28, Anchorage, Alaska, USA, May 2010.
- [11] S. Weiss, M. W. Achtelik, S. Lynen et al., “Monocular vision for long-term micro aerial vehicle state estimation: a compendium,” *Journal of Field Robotics*, vol. 30, no. 5, pp. 803–831, 2013.
- [12] J. Sheng, S. Tano, and S. Jia, “Mobile robot localization and map building based on laser ranging and PTAM,” in *Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA '11)*, pp. 1015–1020, Beijing, China, August 2011.
- [13] C. Forster, M. Pizzoli, and D. Scaramuzza, “SVO: fast semi-direct monocular visual odometry,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '14)*, pp. 15–22, Hong Kong, June 2014.
- [14] J. Engel, J. Sturm, and D. Cremers, “Semi-dense visual odometry for a monocular camera,” in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 1449–1456, Sydney, Australia, December 2013.
- [15] F. Devernay and O. Faugeras, “Straight lines have to be straight,” *Machine Vision and Applications*, vol. 13, no. 1, pp. 14–24, 2001.
- [16] P. J. Huber, *Robust Statistics*, Springer, Berlin, Germany, 2011.
- [17] E. Rosten and T. Drummond, “Machine learning for high-speed corner detection,” in *Computer Vision—ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7–13, 2006. Proceedings, Part I*, vol. 3951 of *Lecture Notes in Computer Science*, pp. 430–443, Springer, Berlin, Germany, 2006.
- [18] J. Martin and J. L. Crowley, “Comparison of correlation techniques,” in *Proceedings of the International Conference on Intelligent Autonomous Systems*, pp. 86–93, Karlsruhe, Germany, 1995.
- [19] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2003.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

