

## Research Article

# Image Antiblurring and Statistic Filter of Feature Space Displacement: Application to Visual Odometry for Outdoor Ground Vehicle

Xiangmo Zhao , Haigen Min, Zhigang Xu , Xia Wu, Xiaoqi Li, and Pengpeng Sun

College of Information Engineering, Chang'an University, Xi'an 710064, China

Correspondence should be addressed to Zhigang Xu; xuzhigang@chd.edu.cn

Received 31 July 2017; Revised 6 November 2017; Accepted 4 December 2017; Published 29 April 2018

Academic Editor: Yunyi Jia

Copyright © 2018 Xiangmo Zhao 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.

Precise, reliable, and low-cost vehicular localization across a continuous spatiotemporal domain is an important problem in the field of outdoor ground vehicles. This paper proposes a visual odometry algorithm, where an ultrarobust and fast feature-matching scheme is combined with an effective antiblurring frame selection strategy. Our method follows the procedure of finding feature correspondences from consecutive frames and minimizing their reprojection error. The blurred image is a great challenge for localization with a sharp turn or fast movement. So we attempt to mitigate the impact of blur with an image singular value decomposition antiblurring algorithm. Moreover, a statistic filter of feature space displacement and circle matching are proposed to screen or prune potential matching features, so as to remove the outliers caused by mismatching. An evaluation of benchmark dataset KITTI and real outdoor data, with blur, low texture, and illumination change, demonstrates that the proposed ego-motion scheme significantly achieved performance with respect to the other state-of-the-art visual odometry approaches to a certain extent.

## 1. Introduction

Precise, reliable, and low-cost vehicular localization across a continuous spatiotemporal domain of highway transportation is an important problem in the field of outdoor ground vehicles. Solving this issue is a breakthrough in promoting the future development of intelligent transportation, specifically in the following five aspects:

*1.1. Conducive to Fine, Real-Time Traffic Information Accurate Perception.* Compared with the traditional video-based or magnetic coil-based detection method, we can obtain more refined real-time traffic parameters by the high-precision vehicle position. With this information, we can calculate accurate micro and macrotraffic parameters (e.g., time headway, space headway, lane occupancy rate, average speed, traffic flow, traffic density, queue length, driving trajectory, congestion level, and accurate OD

(origin-destination)). These parameters provide powerful data support for advanced intelligent transportation applications (e.g., vehicle collision avoidance, parking induction, navigation, platoon collaborative control, and road network dynamic traffic allocation).

*1.2. Conducive to Establishment of the Vehicle Credit Records in the Vehicular Network.* With the long time record of the precise position of the vehicle, the behavior safety level of the vehicle can be evaluated according to the driving trajectory of the vehicle. The recorded credit is valuable to vehicle collision warning, vehicle road right dynamic allocation, vehicle priority ranking, and other applications in the vehicular network.

*1.3. Conducive to Early Warning of Traffic Accidents.* In the future intelligent transportation system, if each vehicle can obtain its own precise position and that of the surrounding vehicle in real time, it can fuse other information (e.g.,

weather, road condition, road line shape, and obstacle) from the roadside to evaluate vehicle risk dynamically. The advance warning of traffic accident can avoid or reduce the probability of traffic accidents.

*1.4. Conducive to Postanalysis of Traffic Accidents.* In the future vehicular network, if the real-time location of the vehicle can be obtained and recorded exactly, microscale data in “before the accident-in the accident-after the accident” scenes can be obtained, which is favorable to reconstructing the situation before the collision and affirming the responsibility scientifically and objectively. The long-term observation of accurate data of traffic accidents is helpful to discovering the distribution law of traffic black spots in time and space, which can provide an important basis for road design and traffic accident prevention.

*1.5. Conducive to the Cooperative Security Decision-Making Analysis with V2V (Vehicle to Vehicle) and V2I (Vehicle to Infrastructure).* A precise and reliable location is the most basic and important data for dynamic traffic information collection and integration and security control with V2V and V2I. The accuracy localization of the vehicle helps to get more refined real-time traffic parameters according to these precise traffic parameters of the vehicle: the optimal speed in the designated area, the probability of joining or leaving a platoon, and the optimal path selection under the premise of comprehensive consideration of trip time and driving distance.

In short, a major breakthrough of large-scale, precise, reliable, low-cost, and real-time vehicular localization will greatly reduce the road traffic system uncertainty and randomness, make the road traffic system further close to the rail transit system, and greatly improve the transport efficiency and safety without reducing the mobility at the same time. In this paper, a precise, reliable, and low-cost ego-motion estimation methodology, especially ones involving an image antiblurring and statistic filter of feature space displacement framework, is proposed. There are three main contributions of this paper.

- (1) An image singular value decomposition (ISVD) antiblurring algorithm is proposed to mitigate the effect of image blur. To highlight the efficiency and performance of the proposed method ISVD, many comparisons have been conducted with other image-blurred degree assessment algorithms.
- (2) For feature displacement, not only adjacent matched frames ( $I_{lp}, I_{lc}$ ) or ( $I_{rp}, I_{rc}$ ) but also the left and right matched frames ( $I_{lp}, I_{rp}$ ) or ( $I_{lc}, I_{rc}$ ) in a stereo model conform to a Laplacian distribution. Based on this discovery, we proposed a statistic filter of a feature space displacement algorithm to remove the outliers in feature correspondences.
- (3) Experiments are conducted on a public benchmark dataset and data captured with our outdoor ground vehicle platform. We analyze the results from quantitative and qualitative aspects to demonstrate the superiority of our proposed method.

## 2. Related Work

This paper focuses on the innovation in the front end of the visual information-based ego-motion estimation method, so here we just review some works in the front-end procedure. The accuracy and reliability of feature detection and matching are the bases of the whole vision-based ego-motion estimation method. To eliminate the error caused by feature detection and matching or tracking, some effective tricks have been developed in the previous relevant literatures. Badino et al. [1] used the whole history of the tracked feature points to compute the motion of the camera, which reduced drift significantly with a negligible additional computational cost. They also discovered that the feature interframe tracking error follows a Laplacian distribution in all three dimensions, while this property was not employed to remove the outliers. The concept of temporal flow [2] restricts the uncertainty to reduce error propagation from matches. A more direct way is careful feature selection and tracking, for example, ORB based [3], KLT based [4], and AKAZE based [5]. With further analysis, we find that the feature displacement statistic for a sequential frame conforms to a Laplacian-like distribution with a zero mean. Badino et al. [1] had explored this distribution while they ignored the ability of this distribution in the matching procedure.

In the feature-based framework, common feature detectors, however, cannot detect enough features from a blurred image, especially for the outdoor ground vehicle in a sharp turn. The blur effect can further degrade the matching accuracy between frames (Figure 1(a); the number of good matches is decreasing from 111 to 15 with a more and more serious blur impact) and detect the inaccurate loop closure (Figure 1(b)). Figures 1(b), b1 and 1(b), b2 show an almost similar scene, and Figure 1(b), b2 is blurring. The similarity between Figures 1(b), b1 and 1(b), b2 is 76%, while that between Figures 1(b), b1 and 1(b), b3 is 92%, which has a great effect on reconstruction and localization performance. When an image is captured when the camera is moving during the exposure time, the one-to-one relationship between the scene points and the image points is broken and a certain number of scene points are projected at a single pixel contributing to the final pixel value. This effect is called motion blur. Motion blur is the most common artifact in images captured by a moving robot mounted with a camera. Usually, there are four types of blur, that is, linear blur, rotation blur, zoom blur, and Gaussian blur, considered as to occur in a moving environment. In order to remove or reduce the harm caused by blurry images, generally, there are two ways used, image deblurring and antiblurring. The former [6] recovers the blurred image by deconvolution, which will sacrifice the feature accuracy and is time consuming. The latter [7–9] classifies the images into blurry and clear images and then eliminates the blurry ones. Lee et al. [10] presented robust data association and drift-free simultaneous localization and mapping (SLAM) with blurred images. The blurred images were recovered by deconvolution. Obviously, this deblurring method was not applicable for outdoor ground vehicle localization. Zhao et al. [11] presented an adaptive blurred image classification algorithm, which was effective



FIGURE 1: The blur impact on feature detection and matching and loop closure detection: (a) the impact on feature detection and matching; (b) the impact on loop closure detection.

but could not be applied to outdoor ground vehicles with feature-matching failure. Pretto et al. [12] proposed a visual odometry (VO) method robust to motion blur based on the improved feature detector SIFT [13]. However, only Gaussian blur is considered in this work; the other three common types of blurring effect are ignored.

In general, there are three ways to estimate visual ego-motion, namely, VO, structure from motion (SFM), and SLAM. VO is devoted to the real-time and accurate estimation of camera movement while SFM focuses on 3-dimensional reconstruction as well as camera pose estimation and usually refines the estimation with bundle adjustment. VO can be thought of as a special case of SFM, whereas SLAM's purposes are for localizing a robot in an unknown environment. More specifically, SLAM recovers the trajectory of the camera and builds a map at the same time. The concept of VO was formally put forward by Nister et al. [14], who realized a real-time VO system. The later research was mostly based on this VO framework. The most successful VO applications were NASA's Mars Exploration Rovers, Spirit and Opportunity [15]. An image pyramid was introduced to help feature tracking in 2011 on Curiosity. Howard [16] implemented a stereo VO with an adopted Harris and Fast feature to ensure real-time performance and employed a feature-matching method [17] to find the correspondences. The aforementioned research laid a good

foundation for the development of VO, but they suffer from a low accuracy. Geiger et al. [18, 19] used a simple Sobel template operator to detect the feature for a real-time system. They tested the VO algorithm on the KITTI benchmark dataset and obtained the positioning result with high efficiency and accuracy. A KITTI vision benchmark suite was established by Andreas (MPI Tübingen), Philip Lenz (KIT), Christoph Stiller (KIT), Raquel Urtasun (University of Toronto), and so on. Bellavia et al. [2] proposed key frame matching and loop closure to build a stereo camera SLAM system. Badino et al. [1] integrated features from multiple frames to improve the accuracy of the motion estimation, and they also combined the historical image information to obtain the refined motion estimation. Mur-Artal et al. [20] proposed a feature-based monocular ORB-SLAM system, which selected the points and keyframes for reconstruction and obtained an excellent performance. ORB-SLAM is the most outstanding work with a feature-based system that runs in real time, in small and large indoor and outdoor environments. Differently from the feature-based method, Engel et al. [21] relied on the photoconsistency of high-contrast pixels, including corners, edges, and high-texture areas. What is more, an illumination invariance method removed the effect of brightness changes. DSO [22] and LSD-SLAM [23] abandoned the feature detection procedure and directly used the actual sensor value light received from a certain direction over a certain



FIGURE 2: Our vehicle platform with stereo camera and DGPS.

time period. Hu and Chen [24] combined a monocular and an inertial measurement unit (IMU) with a visual odometer. They employed trifocal tensor geometry information and a multistate constraint Kalman filter in the algorithm architecture to reduce the time consumption and enhance the accuracy of the algorithm. Besides the fusion applications of visual information and IMU, VO is usually used as a supplementary to the global positioning system (GPS) [25, 26]. Fusion approaches are becoming the mainstream of research about the practical positioning and navigation system. Although there are many multisensor solution schemes for localization, the potential of VO has not been unlocked completely. With the development of deep learning (DL) in computer vision, some researchers have imported DL into localization [27, 28]. Typically, DL is used to detect the loop closure, which is different from the conventional way of bag of words (BoW) [29]. For a more comprehensive and in-depth literature survey of the VO algorithms that have been developed, the reader may refer to [30–34].

### 3. Proposed Method

In this section, the system model and motion parameterization are briefly introduced. Figure 2 shows our experimental platform equipped with a high-resolution stereo camera rig (Basler acA1600 GigE, image size  $1200 \times 800$  pixels) and differential GPS (DGPS, RT2000).

Then, the image antiblurring algorithm based on blurred degree calculation with ISVD is presented. What is more, we propose a statistic filter of feature displacement to obtain the ultrarobust feature correspondences.

**3.1. System Model and Motion Parameterization.** In this paper, we parameterize the motion as the vector  $\mathbf{T}$ .

$$\mathbf{T} = \{r_x, r_y, r_z, t_x, t_y, t_z\}, \quad (1)$$

where  $(r_x, r_y, r_z)$  and  $(t_x, t_y, t_z)$  represent the rotation and translation vectors between the consecutive frames, respectively.

In each procedure of pose estimation,  $\mathbf{T}$  is estimated with  $N$  matched points. For stereo VO, there are four feature

points in each matching step, namely,  $(u_{lc}, v_{lc})$ ,  $(u_{rc}, v_{rc})$ ,  $(u_{lp}, v_{lp})$ , and  $(u_{rp}, v_{rp})$ , where l, r, c, and p represent the left, right, current, and previous frames, respectively.  $(X_{lc}, Y_{lc}, Z_{lc})$ ,  $(X_{rc}, Y_{rc}, Z_{rc})$ ,  $(X_{lp}, Y_{lp}, Z_{lp})$ , and  $(X_{rp}, Y_{rp}, Z_{rp})$  are the 3d coordinates corresponding to the four feature points, as shown in

$$\begin{aligned} X &= \frac{(u - c_u) * b}{d}, \\ Y &= \frac{(v - c_v) * b}{d}, \\ Z &= \frac{f * b}{d}, \end{aligned} \quad (2)$$

where

$$\mathbf{C} = \begin{bmatrix} f & 0 & c_u \\ 0 & f & c_v \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

is the camera intrinsic matrix,  $f$  is the focal length,  $(c_u, c_v)$  is the principal point of the camera,  $b$  is the base length, and  $d$  is the disparity.

During the  $i$ th Gaussian-Newton gradient decent iteration, the motion vector  $\mathbf{T}_i = \{r_{xi}, r_{yi}, r_{zi}, t_{xi}, t_{yi}, t_{zi}\}$  is refined by three random features selected by random sample consensus (RANSAC) [35]. The incremental equation represents

$$\mathbf{J}\mathbf{J}^T \Delta \mathbf{T}' = \mathbf{J}\mathbf{r}, \quad (4)$$

where  $\mathbf{r} = \mathbf{w} * [(u, v) - (\hat{u} - \hat{v})]$  is the residual and  $\Delta \mathbf{T}'$  is the homogeneous form of  $\Delta \mathbf{T}$ . Here,  $\mathbf{w}$  is designed for reducing the error caused by inaccurate camera calibration.  $\mathbf{w} = c_u / |u - c_u|$  is small if the feature point is far away from camera.  $\mathbf{J}$  is the Jacobian matrix represented as

$$\mathbf{J}_{u,v}|_T = \begin{bmatrix} \frac{\partial u}{\partial r_x} & \frac{\partial u}{\partial r_y} & \frac{\partial u}{\partial r_z} & \frac{\partial u}{\partial t_x} & \frac{\partial u}{\partial t_y} & \frac{\partial u}{\partial t_z} \\ \frac{\partial v}{\partial r_x} & \frac{\partial v}{\partial r_y} & \frac{\partial v}{\partial r_z} & \frac{\partial v}{\partial t_x} & \frac{\partial v}{\partial t_y} & \frac{\partial v}{\partial t_z} \end{bmatrix}. \quad (5)$$

If the stopping criterion for iteration meets  $|\mathbf{T}^{(i+1)} - \mathbf{T}^{(i)}| < \varepsilon$ , gradient decent iteration is terminated. Otherwise, gradient decent iteration will continue with  $\mathbf{T}^{(i+1)} = \mathbf{T}^{(i)} + (\mathbf{J}\mathbf{J}^T)^{-1} \mathbf{J}\mathbf{r}$ .

**3.2. ISVD Antiblurring Algorithm.** Image blurring is a tough problem in visual applications, especially for outdoor ground vehicles with a sharp turn. Some efforts have been made for deblurring [6, 36] with a deconvolutional model, which is not calculable in a real-time system, for example, VO and SLAM. Zhao et al. [11] designed an adaptive blurred image classification framework, which selected an image in a relatively slightly blurred degree to calculate the ego-motion. This method can eliminate the blur effect to a certain extent while the image-blurred degree calculation algorithm SIGD

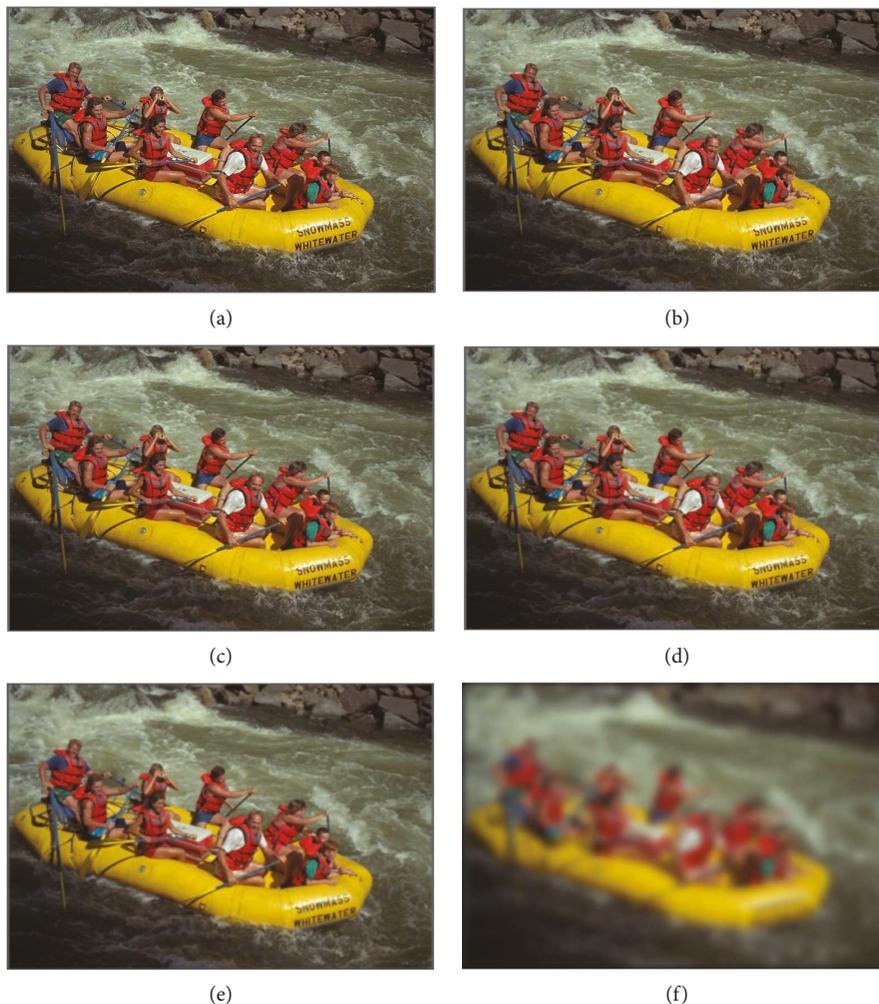


FIGURE 3: Images with different blur degrees from LIVE2 database: image (a): 1.59098; image (b): 9.15433; image (c): 2.37732; image (d): 3.10557; image (e): 4.24759; image (f): 0.84185.

is less impressive than the other no-reference image quality assessment algorithms, for example, Marziliano [7], JNB [8], and CPBD [9]. This paper proposes an outstanding image quality metric assessment method, image singular value decomposition (ISVD), to enhance the performance of image antiblurring.

For an image matrix  $\mathbf{A} \in R^{m \times n}$ , the orthogonal matrixes  $\mathbf{U} \in R^{m \times m}$  and  $\mathbf{V} \in R^{n \times n}$  compose  $\mathbf{A} = \mathbf{U}\mathbf{W}\mathbf{V}^T$ .  $\delta_1 \geq \delta_2 \geq \dots \geq \delta_n$  in  $\mathbf{W} = \text{diag}(\delta_1, \delta_2, \dots, \delta_n)$  are the nonzero singular values. To explore the relationship between the singular value and image quality, we choose six images (Figures 3(a)–3(e)) with an increasing blur degree from the LIVE database [37], as shown in Figure 3.

Figure 4 shows the singular value curve of those different blurred degree images in Figure 3. After decomposing the image matrix, the singular value histogram is built on the nonzero singular values. In this figure, every singular value corresponds to a different singular value number. So we can verify the blur degree with a singular value number without any ambiguity, exactly as the threshold line describes. ISVD is defined as follows:

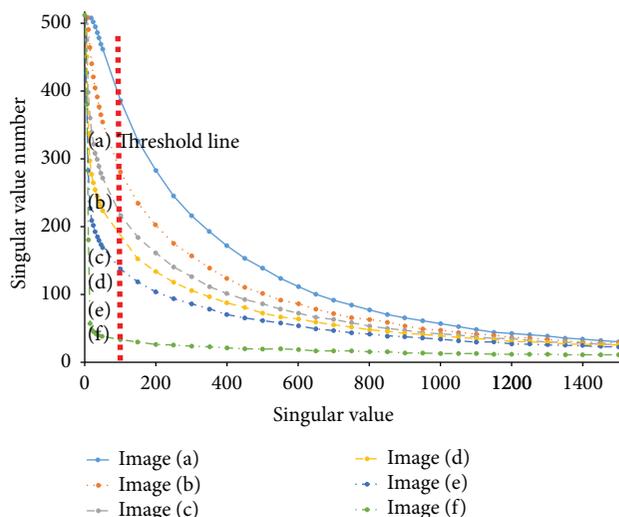


FIGURE 4: The singular value curve for images in Figure 3.

```

Input: image  $I$ .
Output: blurred degree  $b$  of image  $I$ 
Initialization:  $b = 0, num = 0$ 
1  convert  $I$  into grayscale  $I_{gray}$ 
2  calculation singular value  $W = \text{diag}(\delta_1, \delta_2, \dots, \delta_n)$  with singular value decomposition on  $I_{gray}$ 
3  for  $\delta_i$  in  $W$ 
4    if  $\delta_i \geq C_{thres}$ 
5       $num++$ 
6    end
7  end
8  return blurred degree  $b = num/n$ 

```

ALGORITHM 1: Blurred degree calculation with ISVD.

$$\text{ISVD} = \frac{\sum_i^n S}{n}, \quad (6)$$

$$S = \begin{cases} 1, & \text{if } \delta_i \geq C_{thres}, \\ 0, & \text{if } \delta_i < C_{thres}, \end{cases} \quad (7)$$

where  $n$  is the size of a singular value of  $W$  and  $\delta_i$  is the singular value in  $W$ .  $C_{thres}$  is a singular value threshold, and we set 100 in this paper.

The image blurred degree calculation with ISVD is summarized as Algorithm 1.

Then, an adaptive blurred image classification is adopted to select a less blurred image for the following procedures. ISVD's time complexity is  $O(N^3)$ . If the image matrix is too large, the decomposition procedure is too slow. To overcome the disadvantage of time consumption, antiblurring is operated only on keyframe, which performs more efficiently than a frame-by-frame model does.

**3.3. Statistic Filter of Feature Space Displacement.** For the discovery in [9], the feature tracking error follows a Laplacian distribution. We identify that not only the adjacent matched frames ( $I_{lp}, I_{lc}$ ) or ( $I_{rp}, I_{rc}$ ) but also the left and right matched frames ( $I_{lp}, I_{rp}$ ) or ( $I_{lc}, I_{rc}$ ) in the stereo model conform to this distribution. The feature space displacement is the displacement along the  $x$  or  $y$  coordinate in the feature space. In all the original matching features, including inliers and outliers, displacements are collected, the fact being that the displacement for inliers and outliers between frames is fitted with a Laplacian distribution with a long tail, as shown in Figure 5. The corresponding features are more close to the peak bin, which are more likely inliers. Feature displacements for inliers between frames are shown in Figure 6 with a short tail. The  $x$ -axis means the displacement in pixels, and the  $y$ -axis is the number of features collected by histogram statistics.

On the basis of this discovery, we propose a statistic filter of feature space displacement (SFFSD) to remove the outliers and retain the inliers as shown in Algorithm 2. This algorithm only describes the statistic filter between two candidate matched images  $I_1$  and  $I_2$ .

The bin number is set as  $n = 5$ . In that case, the corresponding features cover 80% inliers. However, the value of

$n$  is not set in stone, which depends on the diversity of the detected features. To further refine the matches after statistic filter, circle matching is employed to verify the matches [19]. Three points are randomly selected to implement RANSAC from the above good feature matches at each iteration. The ultrarobust matches improve the accuracy; meanwhile, they accelerate the convergence process of RANSAC.

## 4. Experimental Results and Analysis

**4.1. Performance Evaluation of ISVD.** This paper applies the video quality assessment index proposed by the Video Quality Experts Group (VQEG) to test the performance of different blurred degree assessment algorithms. A few parameters are mentioned as follows. With these parameters, we can evaluate the performance of different algorithms.

- (1) Outlier ratio (OR):

$$\text{OR} = \frac{n_1}{n}, \quad (8)$$

where  $n_1$  is the number of outlier images and  $n$  is the total number of images. If the difference between scores calculated by the blurred degree assessment algorithm and actual reference score is larger than the threshold value, the image is regarded as an outlier image. The smaller the OR is, the better the consistency of the algorithm.

- (2) Root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (s_i - s_{pi})^2}, \quad (9)$$

where  $s_i$  and  $s_{pi}$  are the reference score and score calculated by the blurred degree assessment algorithm of the  $i$ th image, respectively. The smaller the RMSE, the better the predicted result of the assessment algorithm.

- (3) Mean absolute error (MAE):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |s_i - s_{pi}|, \quad (10)$$

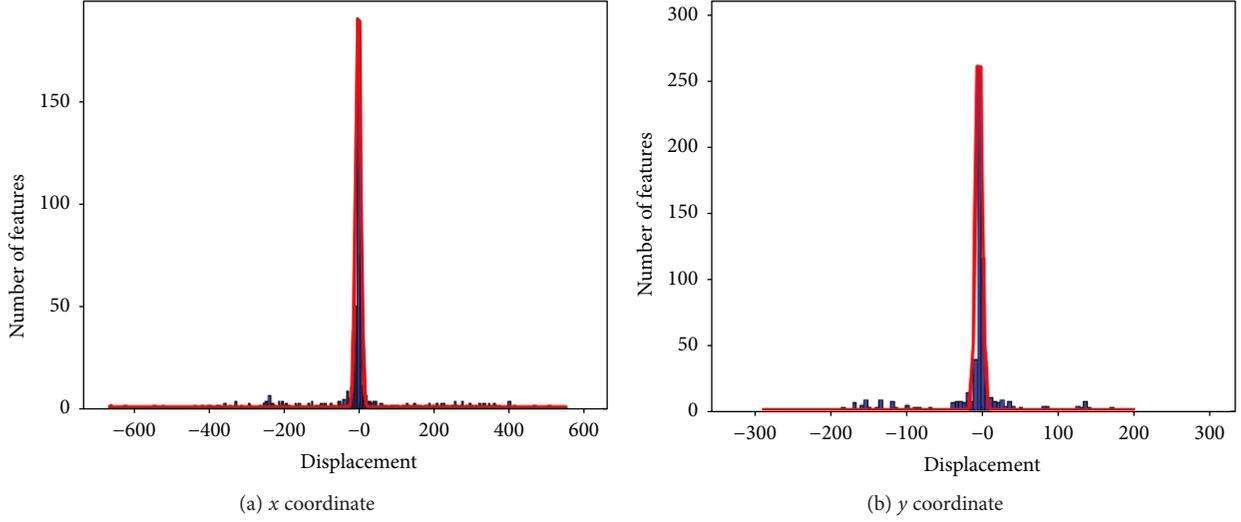


FIGURE 5: Feature displacement between frames, including inliers and outliers. Blue, measured displacement error; red, fitted Laplacian probability distribution function (pdf) with a long tail.

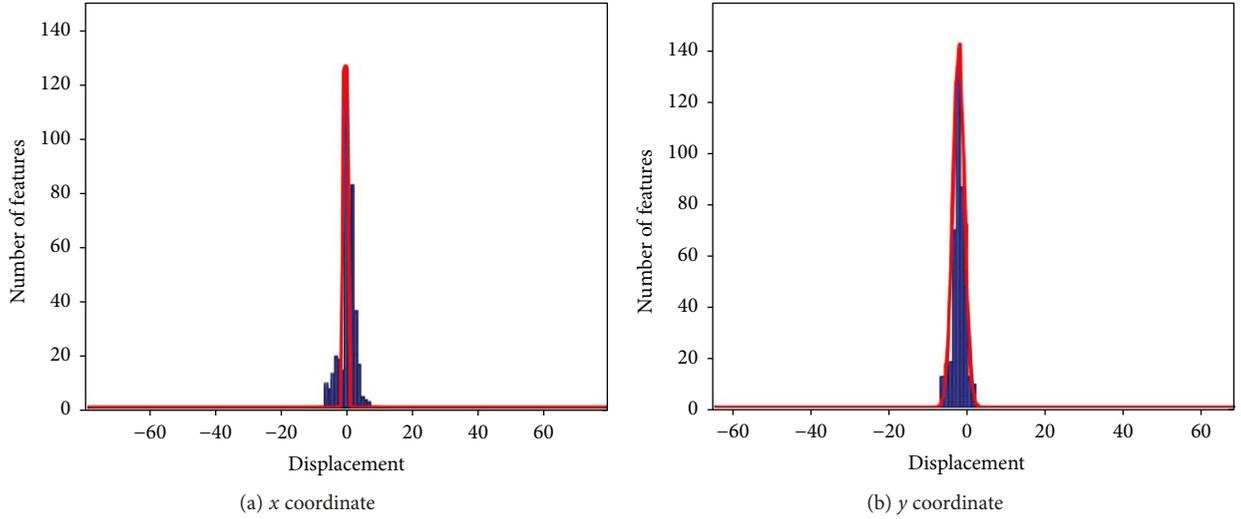


FIGURE 6: Feature displacement for inliers between frames. Blue, measured displacement error; red, fitted Laplacian pdf with a short tail.

where MAE sums the absolute value of the residual  $|s_i - s_{pi}|$ . In statistics, MAE is a measure of the difference between two continuous variables. The smaller MAE is, the better the value calculated by the assessment algorithm.

- (4) Spearman rank-order correlation coefficient (SROCC):

$$\text{SROCC} = 1 - \frac{6 \sum_{i=1}^n (\text{MOS}_{ri} - \text{MOS}_{rpi})^2}{n(n^2 - 1)}, \quad (11)$$

where  $\text{MOS}_{ri}$  (mean opinion scores) is the reference score of the  $i$ th image and  $\text{MOS}_{rpi}$  is the value calculated by the assessment algorithm of the  $i$ th image. The bigger SROCC is, the better the performance of the assessment algorithm.

- (5) Pearson linear correlation coefficient (PCC):

$$\text{PCC} = \frac{\sum_{i=1}^n (\text{MOS}_i - \overline{\text{MOS}})(\text{MOS}_{pi} - \overline{\text{MOS}_p})}{\sqrt{\sum_{i=1}^n (\text{MOS}_i - \overline{\text{MOS}})^2 \sum_{i=1}^n (\text{MOS}_{pi} - \overline{\text{MOS}_p})^2}}, \quad (12)$$

where  $\text{MOS}_i$  is the actual reference score,  $\overline{\text{MOS}}$  is the average reference score of  $n$  images,  $\text{MOS}_{pi}$  is the scores calculated by the assessment algorithm of the  $i$ th image, and  $\overline{\text{MOS}_p}$  is the average calculated score of  $n$  images. The bigger PCC is, the better the consistency between the score calculated by the assessment algorithm and the actual reference score.

With different blurred degree assessment algorithms (Marziliano, JNB, CPBD, SIGD, and ISVD), we calculate the blurred degree for 174 images in dataset LIVE [37]. The actual reference score of those images is known. The image quality parameters OR, RMSE, MAE, SROCC, and PCC are

**Input:** Candidate matched images  $I_1, I_2$ .

**Output:** good feature matches  $VM_{good}$

- 1 detect features  $I_1, I_2$  to get descriptors  $M_{des1}, M_{des2}$  and key-points  $M_{keys1}, M_{keys2}$
- 2 match  $M_{des1}, M_{des2}$  to get the original matches  $VM_{mats}$  with brute force matcher and hamming distance
- 3 Calculate the key-points displacements for  $VM_{mats}$  in x and y components  $\Delta u = u_1 - u_2, \Delta v = v_1 - v_2$
- 4 create a 2D histogram with  $\Delta u$  and  $\Delta v$  to confirm the highest bin for mode approximation
- 5 use the sample within the radius to perform parameter estimation of Laplacian distribution in x and y
- 6 determine the min and max boundary value to include a certain percentage  $ratio = 0.9$  of inliers, assuming a Laplacian distribution
- 7 find the matches  $VM_{u\_mats}$  according to the boundary in Step 6
- 8 repeat Step 6 to Step 7 to find the matches  $VM_{v\_mats}$
- 9 calculate the common element  $V_{com}$  from the  $VM_{u\_mats}$  and  $VM_{v\_mats}$
- 10 for  $id$  in  $VM_{mats}$
- 11     if  $id$  in  $V_{com}$
- 12         pushback the corresponding element into  $VM_{good}$
- 13     end
- 14 end

ALGORITHM 2: Statistic filter of feature displacement.

calculated with different assessment algorithms, as Table 1 shows. Obviously, the proposed ISVD performs more outstandingly than do other algorithms. The blurred degree calculated by ISVD remains monotonic over the full range of the data, as suggested in [36].

The high efficiency is critical to the real-time system. As shown in Table 2, ISVD is less significantly efficient than SIGD, but it is sufficient for real-time VO.

**4.2. Performance Evaluation of Statistic Filter.** We analyze SFFSD's performance against the alternative ratio test (RT; if the ratio of the nearest correspondence distance to the second nearest correspondence distance is less than a threshold, a good match is declared), introduced in SIFT; on the aspect of precision  $P = \text{number of correct matches} / \text{total number of matches}$ , recall  $R = \text{number of correct matches} / \text{number of correspondences}$ , and  $F1$ -measure  $F1 = 2PR / (P + R)$ , which are critical to the matching performance. An evaluation is conducted on Oxford dataset VGG [37, 38] and TUM [39], as shown in Figure 7.

VGG has different photometric and geometric transformations, for example, Gaussian blur, lighting variation, scale, and rotation changes. The first frame in each set is designated as the reference. All other 5 frames, indexed with 0–4, from the same set are matched to the reference. TUM has video sequences with low texture and blurriness. Obviously, SFFSD has relatively higher  $F1$  than the RT.

Figure 8 shows examples of outlier removal with SFFSD, where blue and red lines represent inliers and outliers, respectively. After the SFFSD, circle matching is employed to verify the matches (Figure 9).

**4.3. Experiments and Comparisons.** Experiments have been conducted on the open benchmark dataset KITTI [19] and the real outdoor and indoor datasets. In consideration of feature detection and matching efficiency  $\geq$  efficiency, ORB [37] is regarded as a better choice. The comparison with other detectors and descriptors, like SIFT, SURF [40], KAZE [41],

TABLE 1: Image blur metrics of five different algorithms.

Algorithm	PCC	SROCC	RMSE	MAE	OR
Marziliano	0.8562	0.8672	8.8354	6.4382	0.1264
JNB	0.8498	0.8345	9.0117	6.9317	0.1322
CPBD	0.9257	0.9444	6.4662	4.7098	0.0632
SIGD	0.7596	0.6877	11.1176	9.0331	0.2759
ISVD	<b>0.9405</b>	0.9293	<b>5.8069</b>	<b>4.4090</b>	<b>0.0230</b>

TABLE 2: The execution times of five different algorithms.

Algorithm	Marziliano	JNB	CPBD	SIGD	ISVD
Time (s)	0.5860	0.6156	0.7084	0.0226	0.0082

and AKAZE [42], has been carried out in [41]. To emphasize the excellent performance, we employ two basic VO algorithms, libviso2 (VISO2) [43] and ORB-SLAM2 [3], as comparisons in our experiments. The feature detector and descriptor in VISO2 are replaced with ORB, named ORB-VISO2. Our method is classified into two types, that is, method without loop closure and bundle adjustment based on VISO2 (Our-method1) and method with loop closure and bundle adjustment based on ORB-SLAM2 (Our-method2). No matter what the algorithms Our-method1 and Our-method2 are based on, the major difference is our contribution work of image antiblurring and statistic filter of feature space displacement. Table 3 shows the comparison of processing times among VIOS2, ORB-VISO2, ORB-SLAM2, Our-method1, and Our-method2. The spending time of ISVD and SFFSD is within 13 ms, which has a negligible effect on the total execution time.

**4.3.1. Experiments on Benchmark Datasets.** The first experiment is conducted  $\geq$  conducted on the KITTI benchmark dataset. The KITTI dataset consists of 22 stereo sequences,

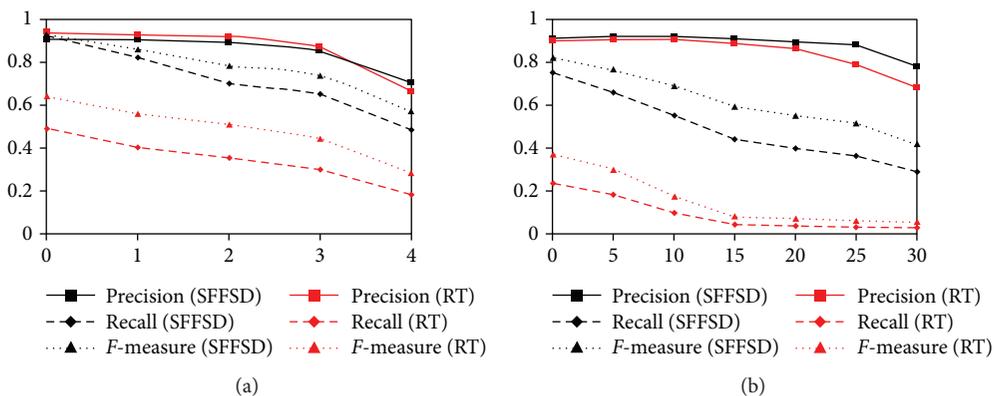


FIGURE 7: Precision, recall, and  $F$ -measure score. The  $x$ -axis of VGG represents the image index, while the  $x$ -axis of TUM represents the relative rotation in degrees. SFPSD is black; RT is red.

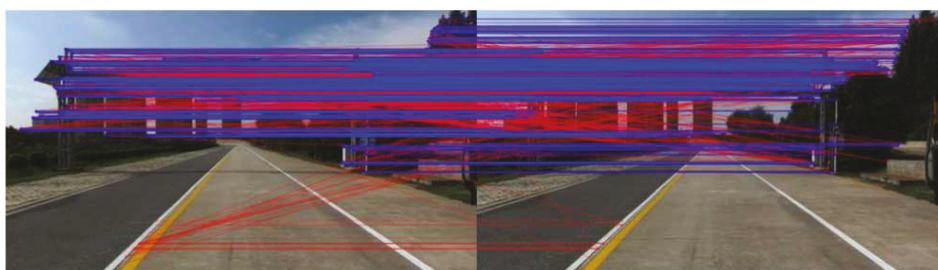


FIGURE 8: Outlier removing with SFPSD. An image pair and its putative correspondences. Blue and red lines represent inliers and outliers, respectively.

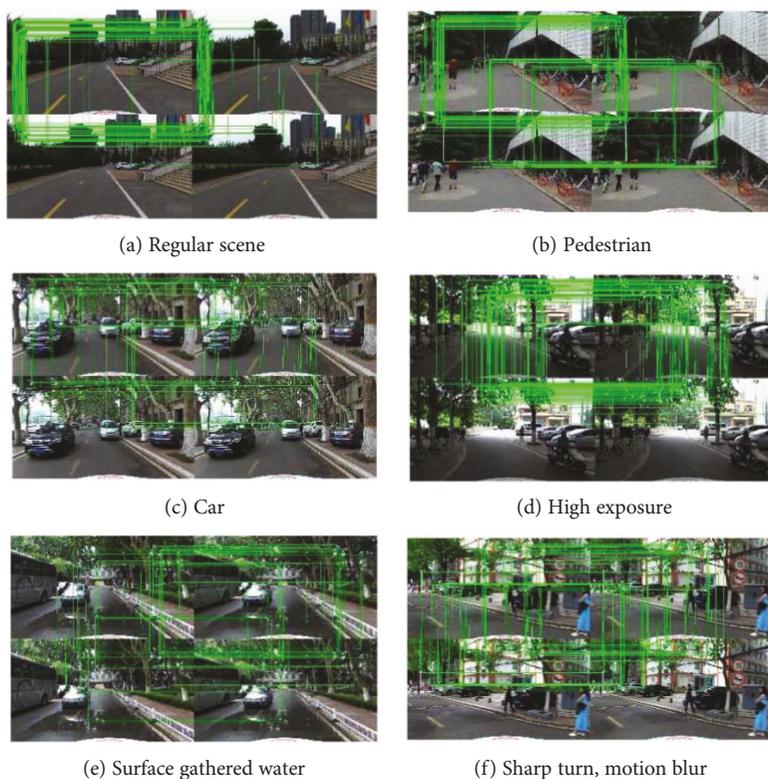
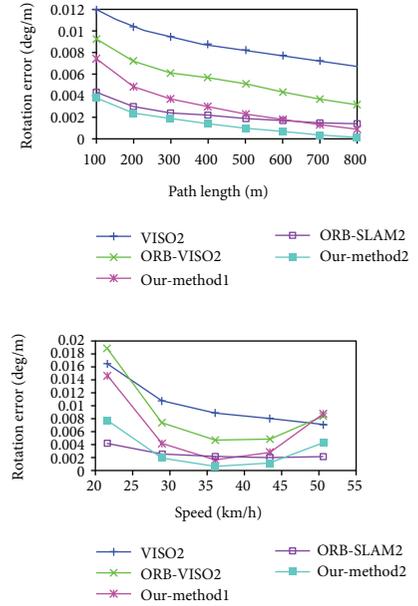
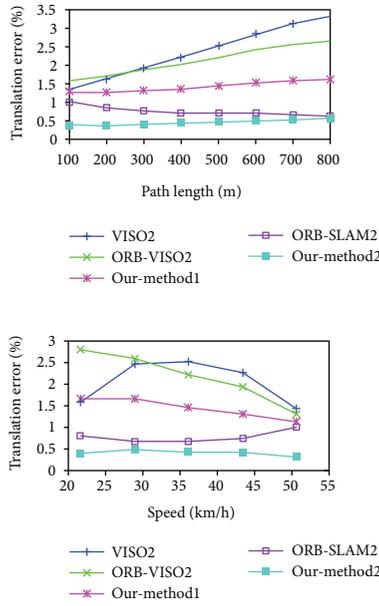
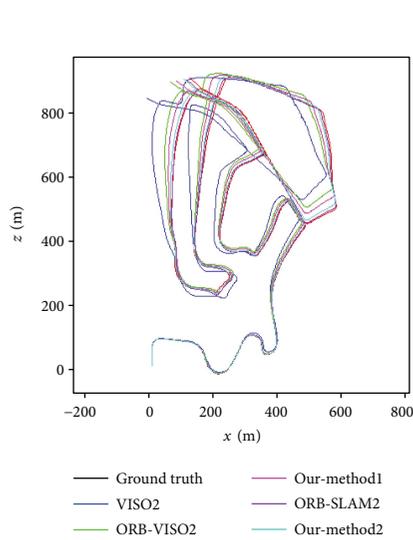


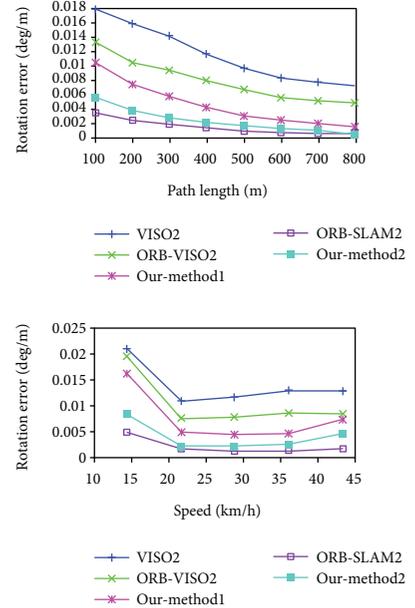
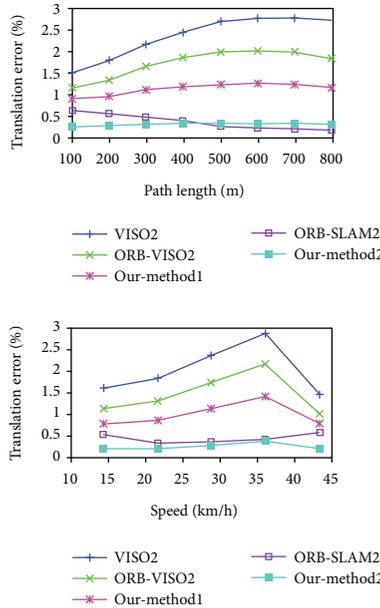
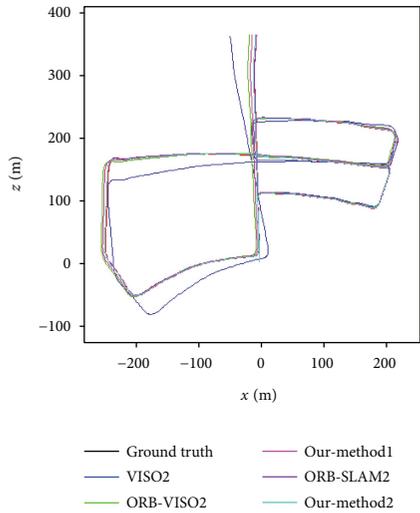
FIGURE 9: The sample images in real outdoor environment and the circle matching described in Section 4.2.

TABLE 3: The average processing time of one image frame (ms).

Method	VISO2	ORB-VISO2	ORB-SLAM2	Our-method1	Our-method2
Time (ms)	17.833	16.130	62.002	21.676	67.519



(a)



(b)

FIGURE 10: Randomly taking sequences 02 and 05 on the KITTI dataset as comparison datasets. Motion estimation on sequences 02 and 05 with ground truth, VISO2, ORB-VISO2, ORB-SLAM2, Our-method1, and Our-method2. The left column shows the trajectory, and the right column shows the translation and rotation error. (a) Comparison result on KITTI 02. (b) Comparison result on KITTI 05.

and we randomly take sequences 02 and 05 as comparison datasets. The trajectories of different algorithms and ground truth are drawn on the same figure.

From Figure 10, we can see that our method is closer to the ground truth than the art-of-the-state method. Regardless of the rotation error or translation error, our method

TABLE 4: Absolute trajectory error RMSE (m).

Sequence	Environment	Distances	VISO2	ORB-VISO2	ORB-SLAM2	Our-method1	Our-method2
KITTI 00	Urban	3714 m	30.133	29.430	1.328	21.016	0.673
KITTI 01	Highway	4268 m	51.430	X	10.936	45.595	4.391
KITTI 02	Urban + country	5075 m	36.441	26.831	5.534	23.185	5.412
KITTI 03	Country	563 m	2.570	2.339	0.619	2.146	0.309
KITTI 04	Country	397 m	1.483	1.3230	0.223	1.190	0.114
KITTI 05	Urban	2223 m	13.653	12.287	0.846	11.740	0.337
KITTI 06	Urban	1239 m	4.318	3.917	0.758	3.757	0.302
KITTI 07	Urban	695 m	5.561	5.017	0.554	4.800	0.221
KITTI 08	Urban + country	3225 m	20.289	X	3.660	17.464	1.470
KITTI 09	Urban + country	1717 m	19.581	16.298	3.427	15.986	1.373
KITTI 10	Urban + country	919 m	4.367	3.812	1.165	3.593	0.468

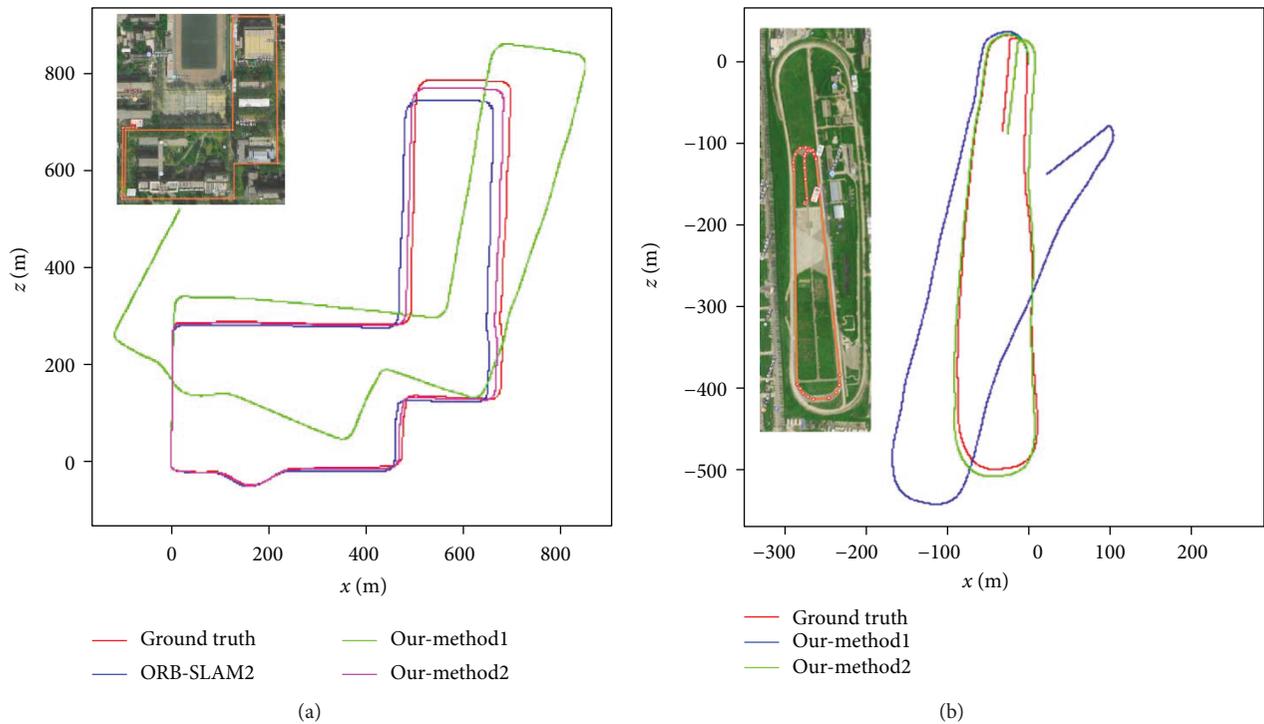


FIGURE 11: Trajectory estimated using the proposed method and DGPS trajectory. The camera frequency was 30 Hz. (a) The trajectory was approximately 2.76 km and lasted 476.4 s. (b) The trajectory was approximately 1.17 km and lasted 109.3 s.

performs better. Compared with other VO algorithms, the absolute trajectory errors of sequence 00–10 are shown in Table 4. In nonloop closure and nonbundle adjustment methods, Our-method1 is better than VISO2 and ORB-VISO2. In sequences “KITTI 01” and “KITTI 08,” ORB-VISO2 fails with fast motion on the highway or illumination changing. In loop closure and bundle adjustment methods, Our-method2 is better than ORB-SLAM2. The longer the trajectory is, the more superiority Our-method2 has.

**4.3.2. Experiments in Outdoor and Indoor Environment.** In this section, to verify the stability and scalability, we also carry out the experiments in real outdoor and indoor environments. The data are captured in a considerably cluttered environment on campus, which contains challenging scenes

like dynamic objects, illumination changing, surface gathered water, and motion blur. The image samples are shown in Figure 9.

The trajectories recovered using the proposed method and other methods are shown in Figure 11(a). The trajectory was approximately 1.45 km and lasted 236.4 s. We regard the trajectory of DGPS as the ground truth. VISO2 and ORB-VISO2 failed to locate with the instability factor in an outdoor environment. It can be found that Our-method1 is not reliable with the error accumulation. ORB-SLAM2 and Our-method2 perform well with loop closure and bundle adjustment. Obviously, Our-method2 has less drift than ORB-SLAM2 has. As shown in Table 5, the proposed method outperforms other methods in terms of overall RMSE and end-point error.

TABLE 5: Overall RMSE and end-point error, corresponding to Figure 11(a).

Error	VISO2	ORB-VISO2	ORB-SLAM2	Our-method1	Our-method2
Overall position RMSE (m)	X	X	9.006	34.961	5.750
Overall orientation RMSE (deg)	X	X	0.108	0.489	0.060
End-point position error (m)	X	X	2.927	87.788	1.175
End-point orientation error (deg)	X	X	0.434	2.506	0.358

TABLE 6: Overall RMSE and end-point error, corresponding to Figure 11(b).

Error	VISO2	ORB-VISO2	ORB-SLAM2	Our-method1	Our-method2
Overall position RMSE (m)	X	X	X	54.961	4.438
Overall orientation RMSE (deg)	X	X	X	0.616	0.079
End-point position error (m)	X	X	X	48.875	1.964
End-point orientation error (deg)	X	X	X	2.464	0.297

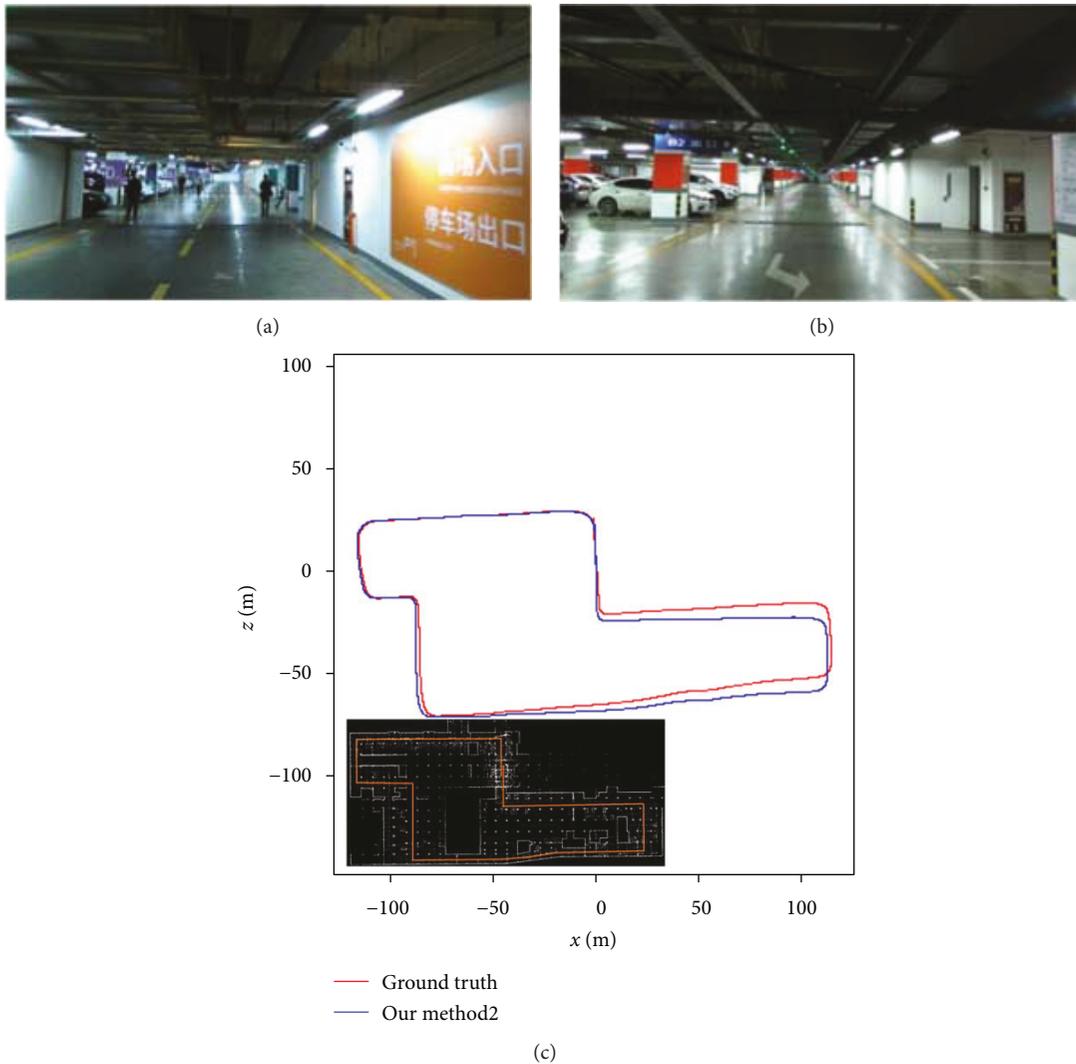


FIGURE 12: Vehicle positioning method results at the parking lot. Different with outdoor experiments, we regard the result of the laser method as ground truth. The map in the left bottom of figure is generated by laser point.

To further support the opinions proposed in this paper, we conduct another experiment to compare the performance of different VO and SLAM kind algorithms. The trajectory in Figure 11(b) is 1.17 km and lasted 109.3 s. The average speed is about 38.5 km/h. Unlike the dataset in Figure 11(a), images in Figure 11(b) contain serious blurring effects or illumination changes, which are common in an outdoor challenging environment. VISO2 and ORB-VISO2, without any exception, fail to estimate the motion of the ground vehicle with this challenging dataset. Because of image blur, moving objects, or illumination change, even ORB-SLAM2 fails to locate the feature. However, Our-method1 and Our-method2 still perform well in the whole process with acceptable rotation and translation accumulative errors. Table 6 shows the overall RMSEs and end-point errors of the listed algorithms corresponding to outdoor data in Figure 11(b).

Figure 12 shows that our proposed methods perform well at the underground parking lot (image samples are shown in Figures 12(a) and 12(b)). Compared with the other vision-based methods, only our proposed method can still locate the vehicle with poor light and dynamic objects while VISO2 and ORB-VISO2 all fail to locate, as shown in Figure 12(c). The repeated indoor parking experiments demonstrate that ORB-SLAM2 is not a stable work because of failure with exquisite lightening change.

## 5. Conclusions

In this paper, an image antiblurring and statistic filter of feature space displacement-based visual odometry algorithm is proposed. The approach achieves a low drift error even for long paths. The impact of image blurring is mitigated by an antiblurring algorithm. A statistic filter of feature space displacement and circle matching are proposed and adopted to screen or prune potential matching features, so as to remove the outliers caused by the mismatching. An evaluation of benchmark dataset KITTI and the real outdoor dataset demonstrates that the proposed method is suitable to be used in robot ego-motion estimation, even for outdoor ground vehicles. Our further research will be devoted to establishing a dense SLAM system based on the proposed method because a dense map is vital to vehicle navigation application.

## Conflicts of Interest

The authors declare that they have no competing interests.

## Authors' Contributions

Xiangmo Zhao and Haigen Min conceived and designed the research. Xiaochi Li, Pengpeng Sun, and Xia Wu performed the experiments. Haigen Min and Zhigang Xu analyzed the data. Xiangmo Zhao, Zhigang Xu, and Haigen Min wrote and edited the manuscript. All authors read and approved the final manuscript.

## Acknowledgments

This work was supported and made possible by the National Natural Science Foundation of China (no. 51278058), 111 Project on Information of Vehicle-Infrastructure Sensing and ITS (no. B14043), Joint Laboratory of Internet of Vehicles sponsored by Ministry of Education and China Mobile (no. 213024170015), Application of Basic Research Project for National Ministry of Transport (Grant no. 2015319812060), and the Fundamental Research Funds for the Central Universities (310824165024, 300102328108, 2013G5240009, and 310824153103).

## References

- [1] H. Badino, A. Yamamoto, and T. Kanade, "Visual odometry by multi-frame feature integration," in *2013 IEEE International Conference on Computer Vision Workshops*, pp. 222–229, Sydney, NSW, Australia, 2013.
- [2] F. Bellavia, F. Bellavia, M. Fanfani, F. Pazzaglia, and C. Colombo, "Robust selective stereo SLAM without loop closure and bundle adjustment," in *Image Analysis and Processing – ICIAP 2013*, pp. 462–471, Springer, Berlin, Heidelberg, 2013.
- [3] R. Mur-Artal and J. D. Tardos, "ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.
- [4] W. Ci and Y. Huang, "A robust method for ego-motion estimation in urban environment using stereo camera," *Sensors*, vol. 16, no. 10, p. 1704, 2016.
- [5] H. Min, X. Zhao, Z. Xu, and L. Zhang, "Robust features and accurate inliers detection framework: application to stereo ego-motion estimation," *KSII Transactions on Internet and Information Systems*, vol. 11, no. 1, pp. 302–320, 2017.
- [6] T. H. Kim and K. M. Lee, "Generalized video deblurring for dynamic scenes," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5426–5434, Boston, MA, USA, 2015.
- [7] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: application to JPEG2000," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 163–172, 2004.
- [8] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 717–728, 2009.
- [9] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2678–2683, 2011.
- [10] H. S. Lee, J. Kwon, and K. M. Lee, "Simultaneous localization, mapping and deblurring," in *2011 International Conference on Computer Vision*, pp. 1203–1210, Barcelona, Spain, 2011.
- [11] H. Zhao, Y. Liu, X. Xie, Y. Liao, and X. Liu, "Filtering based adaptive visual odometry sensor framework robust to blurred images," *Sensors*, vol. 16, no. 7, p. 1040, 2016.
- [12] A. Pretto, E. Menegatti, M. Bennewitz, W. Burgard, and E. Pagello, "A visual odometry framework robust to motion blur," in *2009 IEEE International Conference on Robotics and Automation*, pp. 2250–2257, Kobe, Japan, 2009.
- [13] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International*

- Conference on Computer Vision*, vol. 2, pp. 1150–1157, Kerkyra, Greece, Greece, 1999.
- [14] D. Nister, O. Naroditsky, and J. Bergen, “Visual odometry,” in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*, vol. 1, pp. I-652–I-659, Washington, DC, USA, 2004.
- [15] M. Maimone, Y. Cheng, and L. Matthies, “Two years of visual odometry on the Mars Exploration Rovers,” *Journal of Field Robotics*, vol. 24, no. 3, pp. 169–186, 2007.
- [16] A. Howard, “Real-time stereo visual odometry for autonomous ground vehicles,” in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3946–3952, Nice, France, 2008.
- [17] H. Hirschmuller, P. R. Innocent, and J. M. Garibaldi, “Fast, unconstrained camera motion estimation from stereo without tracking and robust statistics,” in *7th International Conference on Control, Automation, Robotics and Vision, 2002. ICARCV 2002*, vol. 2, pp. 1099–1104, Singapore, Singapore, 2002.
- [18] A. Geiger, J. Ziegler, and C. Stiller, “StereoScan: dense 3d reconstruction in real-time,” in *2011 IEEE Intelligent Vehicles Symposium (IV)*, vol. 32, pp. 963–968, Baden-Baden, Germany, 2011.
- [19] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? The KITTI vision benchmark suite,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3354–3361, Providence, RI, USA, 2012.
- [20] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, “Orb-SLAM: a versatile and accurate monocular slam system,” *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.
- [21] J. Engel, J. Stückler, and D. Cremers, “Large-scale direct SLAM with stereo cameras,” in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1935–1942, Hamburg, Germany, 2015.
- [22] J. Engel, V. Koltun, and D. Cremers, “Direct sparse odometry,” 2016, <https://arxiv.org/abs/1607.02565>.
- [23] J. Engel, T. Schöps, and D. Cremers, “LSD-SLAM: large-scale direct monocular SLAM,” in *European Conference on Computer Vision*, Springer International Publishing, 2014.
- [24] J.-S. Hu and M. Y. Chen, “A sliding-window visual-IMU odometer based on tri-focal tensor geometry,” in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3963–3968, Hong Kong, China, 2014.
- [25] J. Rehder, K. Gupta, S. Nuske, and S. Singh, “Global pose estimation with limited GPS and long range visual odometry,” in *2012 IEEE International Conference on Robotics and Automation*, pp. 627–633, Saint Paul, MN, USA, 2012.
- [26] J. Schneider and W. Forstner, “Real-time accurate geolocalization of a MAV with omnidirectional visual odometry and GPS,” in *International Workshop on Computer Vision in Vehicle Technology*, pp. 483–501, 2014.
- [27] A. Kendall, M. Grimes, and R. Cipolla, “PoseNet: a convolutional network for real-time 6-DOF camera relocalization,” in *2015 IEEE International Conference on Computer Vision (ICCV)*, vol. 31, pp. 2938–2946, Santiago, Chile, 2015.
- [28] X. Gao and T. Zhang, “Unsupervised learning to detect loops using deep neural networks for visual SLAM system,” *Autonomous Robots*, vol. 41, no. 1, pp. 1–18, 2017.
- [29] D. Galvez-López and J. D. Tardos, “Bags of binary words for fast place recognition in image sequences,” *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188–1197, 2012.
- [30] S. Davide and F. Friedrich, “Visual odometry: part I: the first 30 years and fundamentals,” in *IEEE Robotics & Automation Magazine*, 2011.
- [31] F. Fraundorfer and D. Scaramuzza, “Visual odometry : part II: matching, robustness, optimization, and applications,” *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, pp. 78–90, 2012.
- [32] K. Yousif, A. Bab-Hadiashar, and R. Hoseinnezhad, “An overview to visual odometry and visual SLAM: applications to mobile robotics,” *Intelligent Industrial Systems*, vol. 1, no. 4, pp. 289–311, 2015.
- [33] C. Cadena, L. Carlone, H. Carrillo et al., “Past, present, and future of simultaneous localization and mapping: toward the robust-perception age,” *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.
- [34] J. Fuentes-Pacheco, J. Ruiz-Ascencio, and J. M. Rendón-Mancha, “Visual simultaneous localization and mapping: a survey,” *Artificial Intelligence Review*, vol. 43, no. 1, pp. 55–81, 2015.
- [35] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [36] Y. Zhang and K. Hirakawa, “Blind deblurring and denoising of images corrupted by unidirectional object motion blur and sensor noise,” *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4129–4144, 2016.
- [37] Z. Wang, H. R. Sheikh, L. Cormack, and A. C. Bovik, *LIVE Image Quality Assessment Database Release 2* April 2017, <http://live.ece.utexas.edu/research/quality>.
- [38] A. M. Rohaly, J. Libert, P. Corriveau, and A. Webster, “Final report from the video quality experts group on the validation of objective models of video quality assessment,” *ITU-T Standards Contribution COM*, vol. 1, pp. 9–80, 2000.
- [39] K. Mikolajczyk, T. Tuytelaars, C. Schmid et al., “A comparison of affine region detectors,” *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43–72, 2005.
- [40] H. Bay, T. Tuytelaars, and L. V. Gool, “SURF: speeded up robust features,” in *Computer Vision & Image Understanding*, vol. 110, no. 3pp. 404–417, Springer, Berlin, Heidelberg, 2006.
- [41] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, “KAZE features,” in *European Conference on Computer Vision*, pp. 214–227, Springer, Berlin, Heidelberg, 2012.
- [42] P. F. Alcantarilla, “Fast explicit diffusion for accelerated features in nonlinear scale spaces,” in *British Machine Vision Conference*, pp. 13.1–13.11, 2013.
- [43] “LIBVISO2,” May 2016, <http://www.cvlibs.net/software/libviso/>.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

