A Vision-Inertial Odometer Design Based on ORB and Sliding Window

Qinghe Liu\textsuperscript{1,a}, Xue Zhang\textsuperscript{2,b}, Yankun Zhang\textsuperscript{3,c}, Guqi Zhao\textsuperscript{4,d}

\textsuperscript{1}School of Automotive Engineering, Harbin Institute of Technology at Weihai, Weihai, China
\textsuperscript{2}School of Automotive Engineering, Harbin Institute of Technology at Weihai, Weihai, China
\textsuperscript{3}School of Automotive Engineering, Harbin Institute of Technology at Weihai, Weihai, China
\textsuperscript{4}School of Automotive Engineering, Harbin Institute of Technology at Weihai, Weihai, China
\textsuperscript{a}qingsheliu@hitwh.edu.cn, \textsuperscript{b}zhangxuecar@163.com, \textsuperscript{c}20S030124@stu.hit.edu.cn, \textsuperscript{d}zhaogq2020@163.com

Abstract—In order to improve the positioning accuracy and system reliability of the visual odometer in outdoor environment, this paper proposes an unmanned vehicle positioning system using monocular camera and Inertial measurement unit (IMU). The system uses improved ORB algorithm to extract image feature points and establishes a tightly coupled visual-inertial odometer system based on sliding window and nonlinear optimization. Verified by experiment, it can adapt to the environment of changing light, and can complete the positioning task of unmanned vehicles in the outdoor environment.

1. INTRODUCTION
Positioning is one of the key technologies for vehicles to realized autonomous driving. In an open outdoor environment, accurate positioning can be achieved through Global Navigation Satellite System (GNSS), but in some special environments, such as underground parking lots and densely-built blocks, satellite signals are weak and GNSS cannot perform well. Therefore, many unmanned vehicles are beginning to apply vision for positioning and navigation. This kind of system that collects surrounding environment information through the camera and estimates its own motion based on the geometric model of the camera is called the visual odometry (VO). It uses the image information input by the camera to estimate the vehicle's real-time pose, even when the GNSS signal is missing, it can also provide stable positioning information.

However, in the actual scenes of unmanned vehicle, the complicated road environment, extreme lighting conditions and the rapid movement of the camera will lead to large errors in the positioning of the visual odometer. IMU is considered as a good sensor to solve this problem. Under the rapid motion of the camera, the IMU can obtain the angular velocity and acceleration data providing support for pose estimation. Camera can correct the accumulated error of IMU in slow motion. Therefore, the fusion of two sensors can achieve complementary functions.
Visual inertial odometer (VIO) has been gradually implemented in mobile robots in the theoretical verification stage. Therefore, in this study, we have developed a VIO, which combines visual and inertial information to achieve more accurate positioning and navigation of the unmanned vehicle. Experiments show that our improved positioning and navigation algorithm is accurate and robust in unmanned vehicles.

The contributions of this paper include:
- Improved the ORB corner detection algorithm to extract spread evenly feature points.
- Established a VIO model based on sliding window for limiting the complexity of the system and improving the real-time performance.
- Use public data sets and experiment to verify the effectiveness of the algorithm and the feasibility of using it on vehicles.

The rest of this paper is organized as follows: Sec. II analyzes and states the problem of monocular camera and the combine of vision and inertial. Then this paper proposes a tight coupling and nonlinear optimization methods to address this problem and improve the ORB algorithm in Sec. III. Afterwards, experiments are performed to evaluate the proposed methods in Sec. IV. Finally, we conclude our work in Sec. V.

2. RELATED WORK

Studies on monocular visual odometry have made great progress in recent years, such as PTAM[1] and ORB-SLAM[2]. But the common problems faced by all these pure monocular methods are the scale ambiguity of the system and the poor robustness of the algorithm. To solve the problem of monocular system, motion estimation methods using vision and inertial have achieved great attentions.

According to whether the system adds image information to the feature vector, the fusion of visual and inertial can be divided into loosely coupled and tightly coupled. Loose coupling can reduce the complexity of the system, but it ignores the association of different sensor information. So, the accuracy of loosely coupled is not as high as tightly coupled. In recent years, as shown in Tab. 1, some important achievements in visual-inertial SLAM systems have been completed. In the back-end research, the current trend is that more and more research focus on optimization[3].

| Solution name | Publish time | Back-end | Data fusion type |
|---------------|--------------|----------|-----------------|
| MSCKF[4]      | 2007         | Filter   | Tightly coupled |
| ROVIO[5]      | 2015         | Filter   | Tightly coupled |
| OKVIS[6]      | 2015         | Optimization | Tightly coupled |
| VINS-Mono[7]  | 2017         | Optimization | Tightly coupled |
| ORB-SLAM3[8]  | 2020         | Optimization | Tightly coupled |

3. SYSTEM INTRODUCTION

3.1 Framework of System

In order to better adapt to the large-scale scenarios of unmanned vehicles, in our system, we provide a global feature-based SLAM solution based on VINS-Mono, as shown in Fig. 1. This solution enables us to use ORB feature detection algorithm for visual feature extraction. The information obtained by the IMU is pre-integrated to simplify the calculation. Then align the visual and inertial information, and use a tightly coupled method to fuse the visual and inertial information. Use the bundle adjustment (BA) algorithm to optimize the sensor measurement residuals in the sliding window to form a nonlinear optimization visual-inertial odometer system based on the sliding window.
3.2 Extract visual feature points

Due to the real-time requirements of the system and the low computing performance of the computing platform on the vehicle, the improvement of feature point extraction algorithm performance is very important. According to previous research\[9\], compared with SIFT, SURF and other methods, the use of ORB algorithm can meet the real-time requirements. ORB is a corner point extraction algorithm, which improves the non-directional problem of the FAST algorithm, and uses the binary descriptor BRIEF to make the feature matching faster. But ORB does not have scale invariance. In order to solve this problem, as shown in Fig. 2, we use the image pyramid method to convert the image to form different scales, and then extract the ORB corners from the image of each scale. The extracted ORB corners from each scale are aggregated together to form the ORB feature extracted from the image.

Another problem with ORB feature extraction is that the feature points are not evenly distributed on the image. There are often a large number of feature points in some parts, but some parts have no feature points, as shown in Fig. 3. This paper divides the image of each scale into cells. Extract at least 5 feature points on each cell. If 5 feature points cannot be extracted, lower the threshold for extracting feature points. Fig. 4 shows the results of improved ORB feature extraction.

---

**Figure 1.** Framework of the system

**Figure 2.** Image feature extraction pyramid. The original images are down-sampled according to the ratio of 1/1.2 to obtain 8 pictures (including the original image), and then image features are extracted from different scales.

**Figure 3.** The original ORB feature extraction
Feature point matching is the key to data association between two frames of pictures. In this paper, the feature points between the two frames of images are matched according to the Hamming distance calculated by the BRIEF descriptors. In order to reduce the interference of mismatch to pose estimation, the matching results need to be filtered. The steps for feature matching using ORB feature points are as follows:

- Extract FAST corner points from the image.
- Use BRIEF descriptor to describe key corner points.
- Match the feature points according to the Hamming distance of descriptors.
- Filter matching results through RANSAC.

Use improved ORB algorithm to extract the feature of the picture and match it, the effect is shown in the Fig. 5.

3.3 Pure visual initialization

IMU can make up for the lack of scale information in pure visual odometry, but IMU data contains noise and bias. In order to complete the system initialization better, the camera motion is estimated based on the matched ORB feature points. As the monocular camera only knows the 2D pixel coordinates, the camera's pose estimation between two frames can be completed through epipolar constraints as shown in Fig. 6, then using VO to correct the IMU bias and obtain the scale information to complete the system initialization.

![Figure 4. Improved ORB feature extraction](image)

![Figure 5. Improved ORB matching in this paper](image)

![Figure 6. Epipolar constraint](image)
3.4 Nonlinear optimization tightly coupled visual-inertial odometer based on sliding window

After the system is initialized, the fusion of vision and inertial information is completed. Then the camera pose will be estimated through VIO. A sliding window model is built to limit the complexity of the system in this paper. All states information existing in the sliding window is defined as shown below (1):

$$
\chi = [x_0, x_1, \cdots, x_n, x^b, \lambda_0, \lambda_1, \cdots, \lambda_m]
$$

$$
x_k = [p^n_k, v^n_k, q^n_k, b_n, b^g], k \in [0, n]
$$

$$
x^b = [p^b, q^b]
$$

Where $x_k$ is the IMU state when capturing $k^{th}$ frame image. $n$ is the total number of keyframes and $m$ is the total number of feature points in the sliding window. $\lambda_i$ is the inverse depth of the $i^{th}$ feature when it was first observed.

Using the bundle adjustment method to optimize, we minimize the prior estimation and the sum of the Mahalanobis norm of the observation information residuals. In this way, the maximum posterior estimation can be obtained:

$$
\min_{\chi} \left\{ \left\| r_p - H_p \chi \right\| + \sum_{i=1}^{n} \left\| r_p(z^n_{b,i}, \chi) \right\|_{F} + \sum_{i,j\in C} \rho \| r_v(z^v_j, \chi) \|_{F} \right\} 
$$

Where $r_p(z^n_{b,i}, \chi)$ is the residual of IMU measurements. $r_v(z^v_j, \chi)$ is the residual of visual measurements. $B$ is the measured value of all inertial sensors. $C$ is the feature point that has been observed at least twice in the current window. $r_p$ and $H_p$ is the information that needs to be marginalized by the window.

4. EXPERIMENTAL RESULTS

4.1 Vehicle test

The operating environment is configured on the industrial PC, and small-scale operating experiments is conducted with low-cost MYNT camera in outdoor environment. We only use the image from the left camera. The test vehicle is shown in Fig. 7.

![Figure 7. Experimental vehicle equipped with sensors](image)

As shown in Fig. 8, experiments are carried out in outdoor scenarios to verify the pose estimation ability of the algorithm. It can be seen from the experimental results that the system can initially restore the trajectory of the vehicle. Due to this article does not involve loop detection, the system cannot be closed well by observing the start and end points, but this paper presents an idea to improve ORB algorithm and visual-inertial information fusion.
4.2 Public data set testing and analysis

In this section we will use the EuRoC MH_04 public data set to do a simple test on the performance of the algorithm to verify the performance and accuracy of the system.

Figure 8. Trajectory obtained through VIO without loop closing

Figure 9. Trajectory of MH_04, compared with true value.

Figure 10. Comparison of odometer pose and true value in x y z.

In order to evaluate the accuracy of the VIO system proposed in this paper, we compared the true trajectory and estimated trajectory using the MH_04 dataset as shown in Fig. 9 and Fig. 10. The gray line represents the true trajectory, and the red represents the trajectory estimated by the odometer system. Fig. 10 shows the error distribution of the relative pose trajectory of the odometer. The maximum value is 1.256m, the minimum value is 0.0004m, the root mean square error of the whole track is 0.430m.
CONCLUSION

This paper establishes a visual inertial positioning system based on VINS-Mono suitable for unmanned vehicle. In this study, the visual information processing method has been improved. We use a purely visual method to correct the IMU bias and complete the initialization of the system, and establish a VIO system based on a tightly coupled sliding window. Finally, the experiment and simulation have verified that our system is effective.

ACKNOWLEDGMENT

This study is funded by Key R & D plan of Shandong Province (NO.2019GGX104107).

REFERENCES

[1] G. Klein and D. Murray, “Parallel Tracking and Mapping for Small AR Workspaces,” 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, Nara, 2007, pp. 225-234.

[2] R. Mur-Artal, J. M. M. Montiel and J. D. Tardós, “ORB-SLAM: A Versatile and Accurate Monocular SLAM System,” in IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147-1163, Oct. 2015.

[3] H. Strasdat, J. M. M. Montiel and A. J. Davison, “Real-time monocular SLAM: Why filter?,” 2010 IEEE International Conference on Robotics and Automation, Anchorage, AK, 2010, pp. 2657-2664.

[4] A. I. Mourikis and S. I. Roumeliotis, “A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation,” Proceedings 2007 IEEE International Conference on Robotics and Automation, Roma, 2007, pp. 3565-3572.

[5] M. Bloesch, M. Burri, S. Omari, M. Hutter, R. Siegwart. “Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback,” The International Journal of Robotics Research, 2017;36(10):1053-1072.

[6] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, P. Furgale. “Keyframe-based visual–inertial odometry using nonlinear optimization” The International Journal of Robotics Research. 2015;34(3):314-334.

[7] T. Qin, P. Li and S. Shen, “VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator,” in IEEE Transactions on Robotics, vol. 34, no. 4, pp. 1004-1020, Aug. 2018.

[8] C. Campos, R. Elvira, J. J. Gómez Rodriguez, J. M. M. Montiel. “ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM,” arXiv:2007.11898,2020.

[9] E. Rublee, V. Rabaud, K. Konolige and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” 2011 International Conference on Computer Vision, Barcelona, 2011, pp. 2564-2571.

[10] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W.Achtelik, and R. Siegwart, “The euroc micro aerial vehicle datasets,”The International Journal of Robotics Research, 2016;35(10):1157-1163.