CJA 930
13 November 2017
ARTICLE IN PRESS
No. of Pages 13
Chinese Journal of Aeronautics, (2017), xxx(xx): xxx–xxx

Chinese Society of Aeronautics and Astronautics
& Beihang University
Chinese Journal of Aeronautics

cja@buaa.edu.cn
www.sciencedirect.com

Rapid and robust initialization for monocular visual inertial navigation within multi-state Kalman filter

Wei FANG, Lianyu ZHENG *

School of Mechanical Engineering and Automation, Beihang University, Beijing 100083, China

Received 20 September 2016; revised 3 March 2017; accepted 20 June 2017

KEYWORDS

Estimator initialization;
Navigation;
Kalman filter;
Pose estimation;
Visual inertial fusion

Abstract Sensor-fusion based navigation attracts significant attentions for its robustness and accuracy in various applications. To achieve a versatile and efficient state estimation both indoor and outdoor, this paper presents an improved monocular visual inertial navigation architecture within the Multi-State Constraint Kalman Filter (MSCKF). In addition, to alleviate the initialization demands by appending enough stable poses in MSCKF, a rapid and robust Initialization MSCKF (I-MSCKF) navigation method is proposed in the paper. Based on the trifocal tensor and sigma-point filter, the initialization of the integrated navigation can be accomplished within three consecutive visual frames. Thus, the proposed I-MSCKF method can improve the navigation performance when suffered from shocks at the initial stage. Moreover, the sigma-point filter is applied at initial stage to improve the accuracy for state estimation. The state vector generated at initial stage from the proposed method is consistent with MSCKF, and thus a seamless transition can be achieved between the initialization and the subsequent navigation in I-MSCKF. Finally, the experimental results show that the proposed I-MSCKF method can improve the robustness and accuracy for monocular visual inertial navigations.

1. Introduction

Navigation for an isolated target, providing the position, orientation, and velocity of the object in an unprepared environment, is a great challenge in the community of Unmanned Aerial Vehicles (UAVs),[1–3] mobile robot,[4] augmented reality,[5] etc. With an accurate state estimation, the target object can percept the current pose and do the decision-making for the next step, and this technology is becoming ubiquitous in our daily life.

Much work has been done to achieve this assignment, and the visual-based approach is typical one for the pose estimation and the surrounding perception. Liu et al.[6] presented a high-precision pose estimation based on a binocular vision, and the pose can be estimated in real time with the high-speed cameras and laser scanning. As an outside-in capturing style, this method needs to arrange the location and orientation of the observed equipment in advance to avoid the occlu-

* Corresponding author.
E-mail address: lyzheng@buaa.edu.cn (L. ZHENG).

Peer review under responsibility of Editorial Committee of CJA.

ELSEVIER Production and hosting by Elsevier

https://doi.org/10.1016/j.cja.2017.10.011
1000-9361 © 2017 Production and hosting by Elsevier Ltd. on behalf of Chinese Society of Aeronautics and Astronautics.

sion, lacking generality and flexibility for free moving targets' navigation. Zhou and Tao[7] proposed a vision-based navigation method by using marked feature points arranged in a radial shape. Instead of the traditional locating method by PnP, the camera location and outlier points are calculated simultaneously, making the navigation more efficient. Nevertheless, this maker-based navigation method can achieve considerable accuracy only at the expense of the prior knowledge about the scene.

With the progress of the lightweight and cheap micro-electro-mechanical system, the Inertial Measurement Unit (IMU) can provide the position, orientation and velocity passively. However, these consumer-grade IMUs are influenced greatly by the time integration of the systematic error and must be corrected periodically. As a common complementary sensor, the Global Positioning System (GPS) is usually used for this correction,[8] but it cannot work in the GPS-denied environment for an indoor navigation.[9] Thus, Zhu et al.[10] proposed an indoor positioning method by fusing the WiFi and inertial sensors, but WiFi signals are not available everywhere, limiting a wider application of this method to some extent. As described in the last paragraph,[6,7] although there are some limitations by the vision-only navigation method, the accuracy of these pose estimations is reliable. Moreover, the consumer-grade camera and IMU module are ubiquitous in our daily life. For small-size and light-weight applications, such as UAVs and mobile devices, the monocular camera turns out to be a superior visual candidate sensor. Thus, the integrated navigation by combining a monocular camera and an inertial sensor is applied in the paper.

In addition, the visual inertial navigation can be categorized into filtering-based[11] or optimization-based.[12] The filtering-based approach requires fewer computational resources by marginalizing past states, but this method may have slightly lower performance due to the early fixing of linearization points. However, the optimization-based approach results in a better performance at the expense of higher computational demands. The comparison between filtering-based and optimization-based approaches is detailed.[13] Taking the real-time application in the future UAV-like resource-constraint system for consideration, this paper focuses on the filtering-based fusion method with a high efficiency. According to the different combinations of visual and inertial information, the filtering-based sensor fusion solution can be further grouped into loosely-coupled[14,15] and tightly-coupled.[16,17] The loosely-coupled approach usually consists of a standalone vision-based pose estimation module, such as SVO[18] or ORB-SLAM,[19] and a separate IMU propagation module.[15] Due to the lack of cross-relation information between the vision and IMU, the loosely-coupled system is incapable of correcting drifts in the estimator. Tightly-coupled approach performs systematic fusion of the visual and IMU measurements, usually leading to better navigation results.[12] What is more, as a monocular navigation, the arbitrary scale is a common problem, and the tightly-coupled method is able to implicitly incorporate the metric information from IMU into the 3D scene estimation, removing the need for scale modeling individually. Thus, the tightly-coupled method is adapted for our monocular visual inertial navigation.

The most common tightly-coupled visual inertial navigation is to augment the 3D feature positions in the filter state, and concurrently estimates the pose and 3D points.[17] This method leads to an increasing computational complexity with new observed features. To address this problem, the Multi-State Constraint Kalman Filter (MSCKF),[16] instead of estimating the positions of landmarks in the state vector, maintains constant computational demands by a sliding window of camera poses. Compared with the traditional visual inertial navigation in Extended Kalman Filter (EKF), the MSCKF method can balance the efficiency and accuracy within a larger-scale navigation. After the presentation of the MSCKF, a number of improvements have been applied,[20,21] and they increased the accuracy and online calibration performance for MSCKF. Gui and Gu[22] proposed a keyframe within the direct approach in MSCKF, and it can reduce the computational complexity linear to the number of selected patches. However, due to minimum length constraint of sliding windows for state update, the initialization problem in MSCKF is still unsolved. Usually, the sensor module had better be kept static for a while to achieve initialization, and shocks at the initial stage may damage the initialization performance in MSCKF. As a highly nonlinear system by fusion of visual and IMU measurements, an accurate and robust initialization is required to bootstrap the estimator. Hesch et al.[11] had shown that the performance of visual inertial navigation estimator relied heavily on the accuracy of initial state, and the initialization error would be accumulated in the subsequent navigation. In recent years, researchers have made some achievements on the initialization of the visual inertial navigation. Yang and Shen[23] addressed the initialization problem based on optimization-based method. They divided the estimation pipeline into different stages, and the experimental results illustrated the validity of the method. But the optimization-based initialization is of high computational complexity, and the state vector derived from the optimization-based initialization is not consistent with the subsequent MSCKF.

To the best of our knowledge, this is the first paper which pays attention to the initialization within MSCKF framework for visual inertial navigation. In order to improve the initialization performance of MSCKF, a trifocal tensor constraint based Initialization MSCKF (I-MSCKF) method is proposed. Instead of calculating the 3D map points in the scene, the constraints between successive three images are applied to the initial estimator. At the same time, the estimated poses are generated and accumulated to satisfy the minimum sliding window for MSCKF. In addition, in order to improve the initialization accuracy, the sigma-point filter is applied to providing superior approximation for the nonlinear system. Thus, with the proposed I-MSCKF method, the robustness and accuracy of initialization can be improved, resulting in a superior performance for the monocular visual inertial navigation. The overall workflow of the proposed I-MSCKF method is depicted in Fig. 1. The combination of trifocal tensor and sigma-point filter is applied to the initialization, and then the seamless transform for the subsequent state estimation is provided to MSCKF. Finally, the state estimation can be output by the monocular visual inertial navigation.

The following sections are organized as follows. Section 2 introduces the IMU propagation model and the MSCKF method. In Section 3, the trifocal tensor and the sigma-point filter for initialization are depicted. In Section 4, the performance of the proposed I-MSCKF method is demonstrated by different experiments. Finally, the conclusions of the study are summarized in Section 5.

Please cite this article in press as: FANG W, ZHENG L Rapid and robust initialization for monocular visual inertial navigation within multi-state Kalman filter, *Chin J Aeronaut* (2017), https://doi.org/10.1016/j.cja.2017.10.011