Robotics and Autonomous Systems 76 (2016) 1–14

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Metric sensing and control of a quadrotor using a homography-based visual inertial fusion method

Ping Li *, Matthew Garratt, Andrew Lambert, Shanggang Lin

School of Engineering and Information Technology, The University of New South Wales, Australia

HIGHLIGHTS

- The 1-SVD method is shown to be superior over the traditional 2-SVD approach.
- Robustness of the LK algorithm is improved using a transformed binary image.
- A visual inertial fusion method is proposed to estimate metric speed and distance.
- Closed-loop flight proves our approach is suitable for general flight of a MAV.

ARTICLE INFO

Article history:
Received 26 May 2015
Received in revised form
11 November 2015
Accepted 27 November 2015
Available online 4 December 2015

Keywords:
Visual inertial fusion
Optic flow
Homography
Micro Aerial Vehicles

ABSTRACT

The combination of a camera and an Inertial Measurement Unit (IMU) has received much attention for state estimation of Micro Aerial Vehicles (MAVs). In contrast to many map based solutions, this paper focuses on optic flow (OF) based approaches which are much more computationally efficient. The robustness of a popular OF algorithm is improved using a transformed binary image from the intensity image. Aided by the on-board IMU, a homography model is developed in which it is proposed to directly obtain the speed up to an unknown scale factor (the ratio of speed to distance) from the homography matrix without performing Singular Value Decomposition (SVD) afterwards. The RANSAC algorithm is employed for outlier detection. Real images and IMU data recorded from our quadrotor platform show the superiority of the proposed method over traditional approaches that decompose the homography matrix for motion estimation, especially over poorly-textured scenes. Visual outputs are then fused with the inertial measurements using an Extended Kalman Filter (EKF) to estimate metric speed, distance to the scene and also acceleration biases. Flight experiments prove the visual inertial fusion approach is adequate for the closed-loop control of a MAV.

1. Introduction

Rotary-wing Micro Aerial Vehicles (RMAVs), with the capability to hover, take off and land vertically, can operate effectively in low-altitude and cluttered environments. Coupled with their great agility, RMAVs have an advantage over ground robots and fixed-wing MAVs for tasks like bridge or vessel inspection [1] and search and rescue in a partially collapsed building. A comprehensive review for the navigation and control of RMAVs is presented in [2]. Fast, accurate and robust pose estimation is crucial for successful applications of RMAVs. GPS signals are not reliable in a confined space, where RMAVs are often deployed. This paper describes state estimation based on a monocular camera and an Inertial Measurement Unit (IMU). These two sensors are widely available and can be made very compact, energy-efficient and light-weight (can be a few grams [3]).

Many algorithms have been proposed for motion estimation (ME) with cameras, using either feature based approaches or direct methods. The former tries to establish feature (corners, lines or blobs) correspondences between two or more images before computing the camera motion and possibly scene geometry based on a motion model. The latter, as the name suggests, directly solve for the camera motion by minimizing a cost function containing the motion parameters [4]. Feature-based methods, although they only use part of the information in an image, are dominant in the literature because they are easier to implement and more computationally efficient. The two methods are combined to achieve a real-time and high-accuracy visual odometry system in [5].

* Corresponding author.
E-mail addresses: ping.li@student.adfa.edu.au (P. Li), M.Garratt@adfa.edu.au (M. Garratt), A.Lambert@adfa.edu.au (A. Lambert), Shanggang.Lin@student.adfa.edu.au (S. Lin).

http://dx.doi.org/10.1016/j.robot.2015.11.011

The movement of features in two images is often termed as optic flow (OF) and some insects are believed to rely on OF for navigation [6]. Insect-mimicking behaviours have been achieved using OF on robotic platforms, like wall-centring [7], visual homing [8], and obstacle avoidance [9]. OF has also been adopted for hover control [10] and landing [11]. Because they only consider frame-to-frame motion, OF-based techniques are usually subject to long-term positional drift [12]. A snapshot image is captured and acts as a reference for the following images to prevent drifting in hover [13,14]. This method is only effective within the local area where the snapshot image is taken. The same problem exists for methods using artificial patterns [15]. The simultaneous localization and mapping (SLAM) algorithms track features over multiple frames, build a map of these features while at the same time estimating the camera pose. The filter-based SLAM [16] estimates camera pose and feature positions together using an Extended Kalman Filter (EKF) while the keyframe-based SLAM [17] optimizes the feature map using bundle adjustment and recovers camera pose from tracking features of known positions in the map. Both SLAM algorithms have been implemented on RMAVs in [18,19]. A bio-inspired SLAM system, termed as RatSLAM, is also designed for robot navigation [20]. Although attractive, SLAM requires much computation for carefully maintaining a map, detecting loop-closure and performing a large-scale optimization task. In fact, because the complexity of SLAM increases with the scale of the environment, it often reduces to a sliding window odometry approach for constant-time processing [3]. This paper focuses on OF based methods since they can be very lightweight and easy to implement. They can help automatic initialization of a prior map for the monocular SLAM method, and act as a fallback when the latter fails [21]. Specifically, in this paper, a homography model is adopted. The epipolar geometry is independent of the scene structure, however is degenerate over planar scenes or when the camera is purely rotating around the principal axis. Although a homography model assumes planarity of the scene, it is still applicable to scenes where a dominant plane or multiple planes [22] are present. Such scenes are easily found, especially in the man-made world. Using the idea of 'Virtual Parallax, the homography model can be extended to handle non-planar scenes [23]. Normally, after the homography matrix is computed, it is further decomposed to estimate surface normal, rotation and translation (up to scale) with two possible solutions [24]. In this paper, we explain how to compute the homography matrix from the Jacobian motion model [25]. With a new parametrization, it is shown that the unscaled velocity is directly known from the homography matrix, avoiding the need to perform Singular Value Decomposition (SVD) after the homography matrix is calculated. Using real images, we show that our method is more accurate and robust than the method involving the decomposition of homography matrix.

A monocular camera suffers from scale ambiguity. The Parrot AR. Drone scales OF estimation using a downward-facing sonar sensor [26], limiting the operation of the vehicle close to the ground. Sonar sensors may also fail over soft or uneven ground. A scanning laser range finder consumes much power and is cumbersome for a MAV. Whilst the scale can be resolved with another visual sensor, this method requires the two sensors to be placed apart by a proper distance [27,28], making the platform less compact than a monocular system. The method in [28] becomes degenerate for static motion and establishing feature correspondence for a stereo pair is not trivial [27]. In our previous work [29], the scale ambiguity is resolved by fusing the homography based speed estimation (up to a scale) with the acceleration measurements in an EKF framework. The proposed Visual Inertial Fusion (VIF) method is able to estimate metric speed, distance to the terrain and acceleration biases. This paper further examines the OF algorithm on public datasets, provides more details and discussion on the homography model, considers outlier rejection using the popular Random sample consensus (RANSAC) algorithm [30], and evaluates the accuracy of the ME over different textures. Flight tests have also been carried out to prove its effectiveness for closed-loop control of a MAV.

2. Related work

In real-world applications, feature correspondences often contain outliers which could seriously corrupt the solution and thus should be detected and excluded from the estimation. A standard method is the RANSAC algorithm, which computes hypotheses from a randomly generated subset of samples in the dataset and then evaluates the reliability of the results using the rest of the data [30]. The correct solution is chosen as the hypotheses supported by the maximum number of samples. The computational complexity of RANSAC increases dramatically with the minimal number (m) of correspondences required to solve the motion model:

$$itera = \frac{\log(1 - p_s)}{\log(1 - (1 - r_{out})^m)} \tag{1}$$

where $itera$ is the number of required iterations to guarantee a probability of p_s (0.99) for successful estimations; r_{out} means the ratio of outliers in the feature correspondences. To reduce $itera$, a simplified motion model can be chosen so that fewer points are needed (m is smaller). In [31], a 1-point algorithm is developed by taking advantage of the non-holonomic constraints of wheeled vehicles. The restriction of planar motion has led to the development of a 3-point algorithm [32], and even a 1-point RANSAC approach further aided by the relative rotation estimation from the inertial sensors [33]. A 4-point method is proposed, also using relative rotation angle measurements from another sensor [34]. The novelty is that no inter-sensor calibration is required. The Manhattan world assumption is employed to simplify the homography model [35]. Most of these approaches make strict assumptions about the vehicle motion or scene structures and may not be applied to general cases. Another way of decreasing $itera$ is to reduce the outlier ratio by improving the OF algorithm, which is the focus of this paper.

Most OF techniques can be classified as gradient based methods [36] or template matching (TM) based approaches [37]. The former is further divided into local and global schemes [38]. In contrast to the local counterparts, global methods regularize OF through a smoothness term in addition to a data term within an optimization framework. They are very accurate but also computationally demanding. TM has the potential to handle large OF but lacks sub-pixel accuracy and its computational cost scales quadratically with the search range. The Lucas Kanade (LK) algorithm [36] is a local gradient based method and is still widely used since it gives good performance in terms of accuracy, density of motion field and low computational burden [39]. However, like most other OF algorithms, the LK method requires constant brightness of the scene, which is often not the case in reality. To improve the robustness of the intensity based LK, it is proposed in this paper to pre-process the intensity image for extracting information less sensitive to lighting changes. In this way, the main body of the OF algorithm stays unchanged, in contrast to methods that model illumination changes explicitly and needs to solve more parameters [40]. Specifically, the technique in [41] is used that obtains a binary image by comparing the intensities of neighbouring pixels, which has been proven to greatly enhance the robustness of the TM algorithm. This binarization method only requires one comparison per pixel and thus is preferred over other pre-processing approaches that obtain gradient orientation [42] or