Neurocomputing 156 (2015) 105–116

Contents lists available at ScienceDirect

Neurocomputing

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

Fast terrain mapping from low altitude digital imagery

Yawei Luo, Tao Guan*, Benchang Wei, Hailong Pan, Junqing Yu

School of Computer science and technology, Huazhong University of Science and Technology, Wuhan 430074, China

ARTICLE INFO

Article history:
Received 21 October 2014
Received in revised form
22 December 2014
Accepted 29 December 2014
Communicated by Rongrong Ji
Available online 13 January 2015

Keywords:
Terrain reconstruction
Structure from motion
MAV
Aerial triangulation
Large-scale modeling

ABSTRACT

We present a linear time Real Terrain Reconstruction (RTR) framework for fixed-wing micro aerial vehicles (MAVs) in this paper. Single-shot aerial images labeled with GPS and IMU signals are acquired by a fixed-wing MAV in several flights. Then these images are fed into our structure from motion (SfM) processing to generate accuracy pose estimation and 3D points. RTR improves existing state of the art algorithms VisualSFM [1] in multiaspect so as to make it more suitable for large-scale terrain reconstruction from aerial imagery. Firstly, we present a novel strategy of combining signals from airborne sensors (GPS/IMU) with the traditional SfM method, which can improve speed and accuracy of pose estimation observably. Secondly, a delayed aerial triangular method is designed to reconstruct a point visible in more than two cameras with an appropriate baseline. Thirdly, we also release 5 aerial imagery datasets which contain over 15 thousands images totally with the detailed MAV pose information from airborne sensors (GPS/IMU). These resources can be used as a new benchmark to facilitate further research in the area. We test our algorithm on these aerial image sets with various settings, and show that RTR offers state of the art performance for large-scale terrain reconstructions.

1. Introduction

3D reconstruction of real terrain from aerial images has been an active research topic in Computer Vision and photogrammetry in recent years. These research contributed to a significant development of some applications such as Google Earth and Microsoft Virtual Earth, which can deliver effective visualizations of large scale models based on aerial and satellite imagery. Detailed 3D terrains automatically obtained from the real world have been used in many regions such as civil planning, military deployment, virtual tourism, surveillance, path planning, games and movies. Currently, aerial vehicles assembled with active sensors (laser range, scanners, sonar, infrared) have been widely used for these applications. Medium scale aerial vehicles which usually have a lot of payload can carry highly accurate and sophisticated sensors such as long range laser scanners, high performance inertial sensors and a lot of computation power. But for MAVs, which we focus on in this paper, the situation is very different. To use GPS/IMU information straightly in pose estimation is a challenging area of research as the MAV's payload and power restrict it from carrying highly accurate and sophisticated GPS/IMU sensors that can provide high-fidelity pose estimates. Furthermore, GPS signals are highly distorted in urban environment, which lead to that the accurate location estimation purely based on airborne sensors is nearly not possible.

Comparing to the signals from GPS/IMU sensors, traditional SfM technique that is based on the computer vision method is more robust and accurate in small-scale scene. Feature extraction method can typically detect and track hundreds of features in images, which, if properly used, can generate excellent localization results. However, it is difficult to reliably estimate the camera motion over long image sequences because of the accumulated error (drift). At the same time, the high computational complex of the vision method can not be ignored either.

Another problem lies in the small baselines between cameras during triangulation. To generate dense and accuracy models of terrains, aerial images which have large forward overlap and side overlap are acquired by digital airborne cameras. However, these highly redundant images will always result in small baselines, which then cause inaccuracy reconstruction of 3D points. Using the keyframes or skeletal sets [22] filtered from the original images can remit the problem to some extent, but the skeletonization itself is a costly procedure as well.

In this paper, we aim at proposing good strategies to solve the problems discussed above. The contributions which we make in RTR mainly focus on three parts as shown below.

Firstly, we fuse the imprecise GPS/IMU signals with the computer vision method in a unitive framework. Considering that the visual method is precise in the short term and that the GPS/IMU signals are relatively accurate in the long term (see Fig. 1), a novel bundle adjustment taking both advantages is illuminated in

* Corresponding author. Tel: +86 2787556154.
E-mail address: qd_gt@126.com (T. Guan).

http://dx.doi.org/10.1016/j.neucom.2014.12.079
0925-2312/

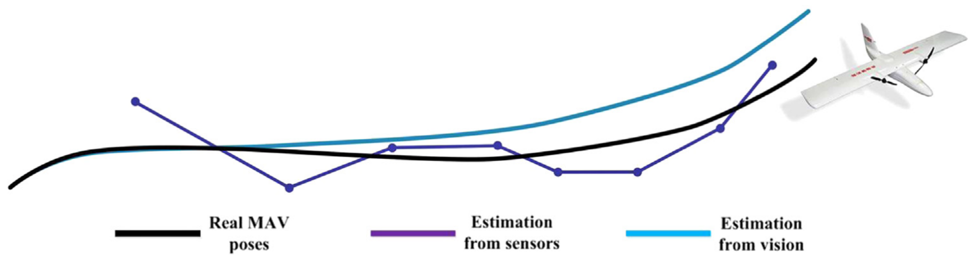

Fig. 1. Different error styles of two methods. Visual method suffers from the accumulative error and the GPS/IMU sensor suffers from accidental error.

Section 4. Furthermore, the hardware signals are expediently used in other aspects in our work. For example, instead of matching all the images to each other, we exploit the approximate GPS tags and match images only to nearby ones. And scale factor is computed from the ratio of the distance between two successive poses of camera and their corresponding GPS in the ground truth to provide geo-location.

Secondly, a novel delayed aerial triangular method that can reconstruct a point visible in more than two cameras with an appropriate baseline is proposed to decrease the error in our SfM framework. Instead of establishing a 3D location using the first joint camera pairs that can observe this point, the new strategy delays the triangulation procedure until another image which has an appropriate baseline with the first camera coming into the iteration. Our method will dynamically determine the number of delays according to the flight height to ensure the best triangulation effect.

Thirdly, we release 5 aerial imagery datasets which contain over 15 thousands images totally with the detailed MAV pose information from airborne sensors (GPS/IMU). These resources can be used as a new benchmark to facilitate further research in the area.

The paper is organized as follows: Section 2 comprehensively reviews the research on both online and offline 3D reconstruction methods on MAVs. Section 3 gives an overview of our RTR system framework. Section 4 discusses our SfM framework we built for aerial imagery in detail, mainly focusing on the delayed aerial triangular method. Section 5 introduces the new fusion bundle adjustment strategy. Section 6 summarizes the steps that generate the final geometrically-accurate dense map of terrain. Then, in Section 7, we show the experimental result and compare it to the recent state of art methods. Finally we conclude the paper in Section 8.

2. Related work

In the last few years there have been significant efforts in the area of large-scale reconstruction, typically targeting natural landscape or urban environment. Here, we discuss research on natural terrain reconstruction from aerial data as it is closely related to our work.

2.1. State of research in online framework

To reconstruct an unknown terrain, a MAV must possess the ability to learn a map of its environment and to localize itself with respect to this map. Many online systems classify the problem as a visual Simultaneous Localization and Mapping (SLAM). Traditional SLAM method is closely related to photogrammetry and computer vision, which tend to address the problem of reconstruction and pose estimation from images only. Though it has been achieved great success using a monocular camera [2–6], it is still difficult to use these methods straightly on some sophisticated platforms such as MAVs. Weiss et al. [18] proposes a vision-only online 3D reconstruction framework on MAV which uses a modified version of the Parallel Tracking and Mapping (PTAM) software for pose estimation with a downward facing camera. But for terrains with large changes in depth (height), only using a mono down-facing camera may not be able to obtain a robust reconstruction. As a result, leveraging additional sensors on MAVs has become main trend in recent years. For instance, Heng et al. [20] presents a real time framework on Quadrotor MAV with an RGBD camera, which also takes advantage of the passive sensors' signals from the MAV to process one-point RANSAC which can decrease the computing complex. Unfortunately, constrained by the sensing range of RGBD camera, the framework is only fit for indoor environment. Furthermore, systems with multi cameras are proposed in [21,22], where the former improves the robustness of the pose estimation by fusing visual odometry readings from a front-looking stereo camera with partial state estimates from a customized downward-looking camera and later relies on the fusion of information from a high frame-rate forward facing fisheye camera, a low frame-rate second camera, and an IMU.

The new problem that how to combine the different sensors smoothly has arisen while it is demanded more robust and accurate through applying additional sensors. Generally, pose estimation combining the different sensors can be regarded as a nonlinear optimization problem. To get the optimize result, filter-based algorithm such as Extended Kalman Filter (EKF) is usually integrity of the online framework [9,10]. Benefiting from the strategy that simultaneously estimates the 3D points and the MAV poses, EKF is more appropriate to be applied in online framework then the iterative-based method such as bundle adjustment (BA). However, EKF-based methods attain lower accuracy than the method of iterative-based and it depends heavily on the effects of initialization as well. In general, the online reconstruction on MAV platform remains as a challenging topic in recent research.

2.2. State of research in offline framework

For the offline reconstruction framework dealing with existed image collections, we prefer to describe it as a SfM method in this paper. SfM methods are very successfully applied in building 3D models from large unstructured collections of images which downloaded from the Internet by using a monocular camera merely [16,17,19]. In order to achieve the “Fast Reconstruction”, [16] leverages massive parallelization technique with the help of parallel compute resources both at CPU level (multi-core) and the network level (cloud computing), but the cost of renting clusters might be high for general users. Snavely et al. [17] constructs skeletal sets of images whose reconstruction approximates the full reconstruction of the whole datasets. However, the graph algorithms themselves can be costly and the number of remaining images can be large. In the meantime, the effects on solution robustness are also not well understood. For SfM system on MAV, fortunately, the monocular camera is also an excellent choice because of its lightweight and power-independent. Warren et al. [19] presents a system for 3D models reconstruction on the MAV using the vision-only method which acts as state of the art. Such results demonstrate the suitability of vision method to large scale pose estimation and mapping tasks on MAV.