ISPRS Journal of Photogrammetry and Remote Sensing 142 (2018) 278–291

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Three-point-based solution for automated motion parameter estimation of a multi-camera indoor mapping system with planar motion constraint

Fangning He*, Ayman Habib

Lyles School of Civil Engineering, Purdue University, 550 W Stadium Ave, West Lafayette, IN 47907, USA

ARTICLE INFO

Keywords:
Motion estimation
Relative orientation
Multi-camera
Structure from motion
Bundle adjustment
Indoor mapping

ABSTRACT

Accurate indoor 3D models have become a key prerequisite for various applications. Through state-of-the-art image processing techniques, 3D models can be generated from high quality images captured by off-the-shelf digital cameras. To acquire redundant data and produce real scale models, a multi-camera system can be used. However, dedicated approaches for image-based 3D reconstruction using mapping platforms equipped with multiple cameras have not been fully addressed. Assuming the availability of prior information regarding the platform trajectory, this paper presents a new approach for reliable estimation of system motion parameters between different data acquisition epochs of a multi-camera system. This approach, which assumes planar motion of the utilized platform, provides a three-point closed-form solution. The derived solutions are then incorporated within a modified RANSAC framework for outlier detection/removal. It is worth noting that, different from the existing General Camera Model (GCM)-based solutions, the proposed approach is based on a modified co-planarity model, which is essentially a direct extension of the classic stereo-based relative orientation. Moreover, since the proposed approach only provides a maximum number of four possible solutions for system motion parameters over different epochs, it has better computational efficiency when compared to other existing algorithms. Experimental results from real datasets acquired with different configurations have demonstrated the reliability of the proposed approach in motion parameter estimation for indoor multi-camera mapping systems.

1. Introduction

3D modelling/reconstruction of indoor environments has been a highly active research area over the past few decades. Accurate 3D models, which contain necessary geometric information of the indoor environments, can be valuable assets to a variety of applications, such as facility management, virtual reality, indoor Geographic Information System (GIS), and Building Information Modelling (BIM). Similar to outdoor mapping procedures, 3D reconstruction/representation of objects in indoor environments can be achieved through either active or passive remote sensing systems. Active sensors, such as laser scanners, are able to directly provide precise and reliable 3D information about scanned objects. However, laser scanning systems are usually expensive, cannot provide sufficient positional information along breaklines, and require direct geo-referencing information especially when collecting data from mobile platforms. The other alternative to generate accurate and textured 3D models for indoor mapping applications is using passive sensors, or simply digital frame cameras. Compared to laser scanners, imaging sensors are more attainable. Nowadays, off-the-shelf digital cameras are capable of acquiring high quality images in different conditions, which is suitable for accurate 3D modelling. However, extensive processing, such as image matching and pose estimation, is always required to extract 3D information from 2D images. Due to some financial and technical constraints, digital frame cameras are still the optimum option for most 3D applications (Remondino and El-Hakim, 2006).

Image-based 3D reconstruction is conventionally established through geo-spatial data acquisition systems that are equipped with a single camera. Thanks to remarkable developments in both multi-sensor integration and synchronization, utilization of multiple cameras has emerged as a popular option for 3D reconstruction to satisfy the needs of a variety of mapping applications. Compared to single camera-based systems, the advantages of using multiple cameras include larger field of view, redundant captured images, and possible true scale 3D modelling by considering the system mounting parameters of each camera relative to a platform body frame or a reference camera (Mazaheri, 2015). So far, several multi-camera systems, such as Point Grey Research/FILR Ladybug (FLIR Systems, 2018), Digital Domain Kronos and

* Corresponding author.
E-mail addresses: he270@purdue.edu (F. He), ahabib@purdue.edu (A. Habib).

https://doi.org/10.1016/j.isprsjprs.2018.06.011
Received 12 January 2018; Received in revised form 14 June 2018; Accepted 18 June 2018

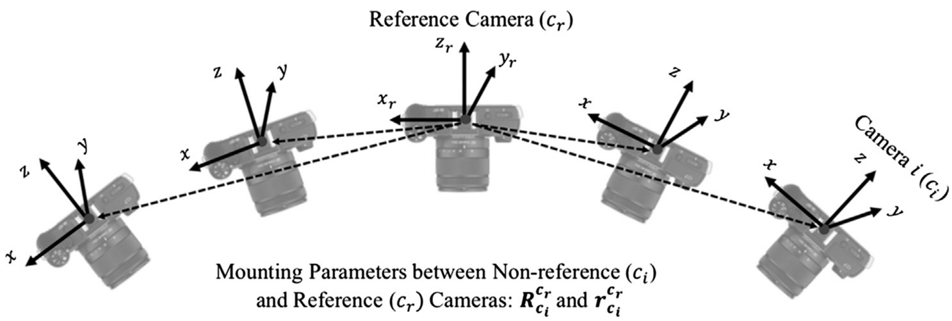

Fig. 1. Reference and non-reference cameras for a multi-camera system with overlapping/converging views.

Zeus (Immersive Media, 2018), have been developed and introduced for different mobile mapping platforms. It is worth noting that since these multi-camera systems are mainly designed for spherical and panoramic image production, the configuration of cameras only ensures minimum overlap among images captured at the same epoch. Due to some financial and technical constraints (e.g., initial investment, mobilization cost, and the required technical expertise of end users), the widespread adoption of these systems has been prevented. In terms of indoor-mapping/close-range applications, a multi-camera system with overlapping/converging views is usually utilized. Meanwhile, in order to provide more affordable image-based 3D reconstruction techniques, a low-cost integrated system, which is usually equipped with multiple off-the-shelf digital cameras, is always preferred. The existing body of literature (Fritsch et al., 2012; Detchev et al., 2011, 2013; Mazaheri and Habib, 2015) has demonstrated the feasibility of deriving accurate 3D models using such low-cost multi-camera systems. Fig. 1 illustrates a commonly-used configuration for a multi-camera mapping system with overlapping/converging views. As shown in the figure, one of the involved cameras is selected as the reference camera (i.e., c_r), and the remaining cameras (e.g., non-reference cameras c_i) are defined relative to the reference one through system mounting parameters (i.e., rotation $R_{c_i}^{c_r}$ and translation $r_{c_i}^{c_r}$). Based on such configuration, the system mounting parameters can be directly incorporated into a set of modified collinearity equations (Habib et al., 2011), which leads to a simpler model that is not affected by the number of utilized cameras or involved epochs. Another advantage of using this configuration is the capability of dealing with multi-camera systems, which are either equipped with a direct geo-referencing component or indirectly geo-referenced (i.e., when GNSS/INS information is not available) (Rau et al., 2011). To date, Structure from Motion (SfM) (Hartley and Zisserman, 2003), which was mainly initiated by the computer vision research community, has been widely adopted for indoor mapping while using multi-camera systems in the absence of GNSS/INS information. Similar to 3D reconstruction using images captured by a single camera, current SfM-based approaches for a multi-camera system are usually implemented in three steps. In the first step, initial feature correspondences are identified in overlapping images. Then, given the corresponding features, the Exterior Orientation Parameters (EOPs) of the reference camera throughout the data acquisition campaign are recovered (the EOPs of the non-reference cameras are defined by the EOPs of the reference camera and the system mounting parameters). Finally, a bundle adjustment (BA) procedure is conducted to refine the derived image EOPs and object coordinates. It is worth noting that bundle adjustment can achieve the best 3D reconstruction accuracy, provided that sufficiently accurate approximations of the unknowns and correct feature correspondences are utilized. In this regard, one can conclude that accurate estimation of the system motion parameters, which define the EOPs of the reference camera at each imaging epoch, is an important prerequisite for the multi-camera based 3D reconstruction.

In the past few decades, motion parameter estimation for multi-camera systems has been investigated within both photogrammetric and computer vision research communities. Similar to relative orientation recovery for single camera-based systems, motion parameter estimation of multi-camera systems relies on reliable conjugate point pairs in overlapping images. In practice, due to the increasing ambiguity arising from repetitive patterns or insufficient texture, identified feature correspondences among overlapping images captured in indoor environments are usually contaminated by a high percentage of outliers. In order to mitigate the impact of outliers, approaches for motion parameter estimation, which take advantage of prior information regarding the trajectory of the utilized platform, have been developed and utilized for reliable 3D reconstruction using either single or multiple cameras (He and Habib, 2016). In this paper, we introduce a new approach for motion parameter estimation between successive imaging epochs that are captured by a multi-camera indoor mapping system with planar motion constraint. The proposed approach has the following characteristics:

- It assumes that the involved images are acquired by a converging multi-camera system with a planar motion constraint, which leads to a three-point closed-form solution for the system motion parameters between successive epochs.
- It is based on a modified co-planarity model, which describes the epipolar geometry within overlapping stereo-images between successive epochs while considering the rigid mounting parameters among different cameras. The utilization of such modified co-planarity model, which is essentially a direct extension of stereo-based relative orientation, simplifies system motion estimation to the recovery of a modified 3×3 essential matrix with an absolute scale.
- It is coupled with a modified RANSAC process for outlier removal. Therefore, it can deal with sets of images where the initial point correspondences are contaminated by a high percentage of outliers.

The remainder of this paper starts with a literature review of related work, which is followed by the mathematical details of the proposed approach for automated motion parameter estimation. Afterwards, experimental results using real datasets are presented. Finally, conclusions and recommendations for future work are introduced.

2. Related work

To date, several approaches, which are mainly motivated by the concept of Generalized Camera Model (GCM), have been introduced for deriving reliable estimates of system motion parameters for multi-camera systems. The GCM, which describes the relative orientation of multiple cameras mounted rigidly on a single body over two different epochs, is first introduced by Grossberg and Nayar (2001) and then further explained by Pless (2003). The GCM is a model for an imaging situation, where pixels in the image correspond to specified light rays (straight lines) in space, but with no other limitation on how incoming light rays project onto an image (Kim et al., 2010). The main difference between the GCM and the widely-adopted pinhole camera model is the absence of a single center of projection (Lee et al., 2013a). In practice, the GCM can be incorporated with different types of imaging systems