Metric Calibration of Aerial On-Board Multiple Non-overlapping Cameras Based on Visual and Inertial Measurement Data

Xiaoqiang Zhang[1](✉), Liangtao Zhong[1], Chao Liang[1], Hongyu Chu[1], Yanhua Shao[1], and Lingyan Ran[2]

[1] School of Information Engineering, Southwest University of Science and Technology, Mianyang, Sichuan, China
{xqzhang,chuhongyu,}@swust.edu.cn, zhongliangtao@mails.swust.edu.cn, syh@cqu.edu.cn

[2] School of Computer Science, Northwestern Polytechnical University, Xi'an, Shaanxi, China
lran@nwpu.edu.cn

Abstract. Recently, the on-board cameras of the unmanned aerial vehicles are widely used for remote sensing and active visual surveillance. Compared to a conventional single aerial on-board camera, the multi-camera system with limited or non-overlapping field of views (FoVs) could make full use the FoVs and would therefore capture more visual information simultaneously, benefiting various aerial vision applications. However, the lack of common FoVs makes it difficult to adopt conventional calibration approaches. In this paper, a metric calibration method for aerial on-board multiple non-overlapping cameras is proposed. Firstly, based on the visual consistency of a static scene, pixel correspondence among different frames obtained from the moving non-overlapping cameras are established and are utilized to estimate the relative poses via structure from motion. The extrinsic parameters of non-overlapping cameras is then computed up to an unknown scale. Secondly, by aligning the linear acceleration differentiated from visual estimated poses and that obtained from inertial measurements, the metric scale factor is estimated. Neither checkerboard nor calibration pattern is needed for the proposed method. Experiments of real aerial and industrial on-board non-overlapping cameras calibrations are conducted. The average rotational error is less than 0.2°, the average translational error is less than 0.015 m, which shows the accuracy of the proposed approach.

Keywords: Non-overlapping field of view · Metric calibration · Camera-imu system · Unmanned aerial vehicle

Supported in part by the National Natural Science Foundation of China under Grant 61902322, in part by the Doctoral Fund of Southwest University of Science and Technology under Grant 19zx7123, in part by the Open Fund of CAUC Key Laboratory of Civil Aviation Aircraft Airworthiness Certification Technology under Grant SH2020112706.

H. Ma et al. (Eds.): PRCV 2021, LNCS 13020, pp. 16–28, 2021.
https://doi.org/10.1007/978-3-030-88007-1_2

1 Introduction

With the continuous development of unmanned aerial vehicle (UAV) and aerial photography technology, various aerial vision applications have been developed. The mobility and the ability of active surveillance make the UAV a good platform for remote sensing, object detection and tracking. In applications like visual surveillance, a larger FoV of the on-board camera would enhance the performance of the algorithm. Monocular large FoV cameras such as the fish-eye camera usually have severe distortions, which may change the appearance of the object and hence affect the performance in object detection or tracking. Recently, multi-camera system with limited or non-overlapping FoVs are becoming popular in the virtual reality community [5]. Compared to a monocular system, multiple non-overlapping cameras would have more FoVs. Besides, fewer cameras would be used in such a system compared to the conventional multi-view system with common FoVs. The reduced weight and power consumption of the multiple non-overlapping cameras system make it suitable for on-board equipment for the UAV. Figure 1(a) illustrates a self-build aerial on-board non-overlapping camera system with 4 cameras. With a carefully designed 3D-printed bracket, the entire system can be easily mounted on a UAV (red dashed bounding box in Fig. 1(a)). The total weight of the system is around 160 grams. Figure 1(b) gives the images captured from the system. It can be seen in Fig. 1 that there is no common FoV between adjacent cameras in the system.

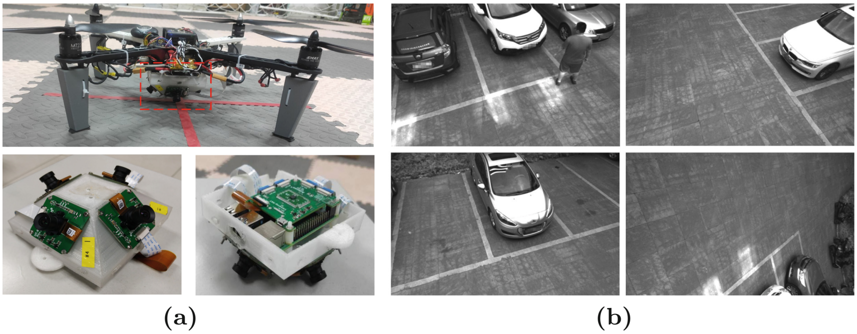


Fig. 1. The aerial on-board non-overlapping camera system and its captured images. **(a)** A self-build non-overlapping camera system on a UAV. **(b)** Images of a parking lot captured from the non-overlapping camera system (Color figure online)

In many vision applications, accurate calibration of the visual system is important to the performance of algorithms. Conventional multi-view system can either be extrinsically calibrated from Zhang's method [23] or from self calibration method [20] using calibration object in the common FoV. However, for a system with multiple non-overlapping cameras, such approaches are difficult to adopt due to the non-overlapping FoVs. In order to calibrate the relative poses of the non-overlapping cameras, different approaches are proposed to establish the pixel correspondences among images with limited or no common FoVs. Some of these approaches would rely extra cooperative calibration objects like a mirror [10,19], multiple checkerboards [9,12,18,22], or designed planar patterns [7,8,11,21]. Other methods are usually based on moving camera pose estimation techniques like structure from motion [17]. However, moving camera based approaches [14,15,24] usually suffers from the lack of metric information, i.e., the physical distance in the real world. The scale ambiguity motivates us to seek a metric calibration approach for multiple non-overlapping cameras.

In this paper, we intend to solve the metric calibration problem of the UAV on-board multiple non-overlapping cameras, that is, the estimation of the relative rotations and translations in metric scale between each cameras and one selected reference camera in the system. A novel visual and inertial data based multiple non-overlapping cameras metric calibration is proposed. By jointly moving the entire system and capturing image sequence of a static scene, the correspondences between 2D pixel locations of feature points and 3D scene points are established. The camera pose of each image can be estimated via solving the perspective-n-point problem. To solve the scale ambiguity problem in the estimated poses, we make use the UAV on-broad inertial measurement unit (IMU) to estimate the metric scale factor. Real experimental results of two different multiple non-overlapping cameras systems quantitatively demonstrates the accuracy of the proposed calibration method, which are comparable with the multiple checkerboards based approaches. The metric scale estimation accuracy is further quantitatively verified in object 3D reconstruction experiments.

2 Related Works

In the literature, the calibration for multiple camera system with limited or no common FoVs can be roughly grouped into two categories, namely cooperative calibration objects based methods and moving cameras based methods.

Early work in [9] utilizes a 3D object with known geometry to estimate the relative poses among multiple cameras. Compared to the carefully designed and fine 3D objects, more commonly used objects in the multiple non-overlapping cameras calibration procedures are the planar checkerboards [9,12,18,22] or designed patterns [7,8,11,21]. Liu et al. [12] propose a method to calibrate the non-overlapping cameras using a compound object consisting of two unknown geometry of checkerboards. Yin et al. [22] introduce an extrinsic calibration method of non-overlapping cameras by solving linear independent equations. By moving the multi-camera system at least twice, enough constraints can be established. The rotations and translations are optimized separately. For pattern based approaches, Li et al. [11] propose a feature descriptor-based calibration pattern, which contains more features of varying scales than checkerboards. Pattens can be recognized and localized even if the pattern is partially observed. Xing et al. [21] design a patten with different types of identity tags or textures in the blank regions of a checkerboard, which also allows calibrations with partially observed patterns. To overcome the problem of non-overlapping FoVs, mirrors are also used for calibration [10,19], which allows cameras to observe the calibration object via reflection. However, sophisticated light path designs are usually needed for these mirror based systems, which restricts the flexibility of the calibration.

Multi-camera calibration based on moving camera method are inspired by Gaspi and Irani [6], which assumes that the lack of common FoV can be compensated by the movements of cameras. In the literature, structure from motion (SfM) has been used to calibrate multiple camera systems [14,15,24]. Zhu et al. [24] extends the single-pair hand-eye calibration used in robotics to multi-camera systems. By utilizing the planar structures in the scene, a plane-SfM is proposed for multiple non-overlapping cameras. High-precision measuring device or specially designed calibration objects are not needed in these approaches. However, because the decomposition of epipolar matrix can only estimate the relative translations between two cameras up to an unknown scale, SfM based approaches usually suffers a scale ambiguity.

3 Metric Calibration Based on Visual and Inertial Measurement Data

3.1 Notation and Problem Formulation

To formally describe the proposed approach, we first introduce coordinate systems utilized in our work. Suppose that n cameras $C_1, C_2, \ldots, C_n$ and an IMU are rigidly connected in the non-overlapping camera system. During the calibration procedure, the system is jointly moving. $C_{i,t}$ denotes the local camera coordinate system of camera C_i at time t, I_t is the inertial coordinate system at time t. We use W to denote the world coordinate system of a static 3D scene.

A notation system of superscripts and subscripts is used for denoting vectors, the relative rotation and translation between different coordinate systems. Vectors in coordinate system A is denoted by the superscripts, and a time varying vector is denoted by a subscripts t. For example, $\mathbf{P}^A$ is the coordinates of a 3D point $\mathbf{P}$ in coordinate system A, $\mathbf{v}_t^A$ is a time varying vector in A. Suppose $\mathbf{P}^A, \mathbf{P}^B$ are the coordinates of a 3D point $\mathbf{P}$ in coordinate system A and B, respectively, we would have $\mathbf{P}^B = R_A^B \mathbf{P}^A + \mathbf{t}_A^B$. $R_A^B, \mathbf{t}_A^B$ denote the relative rotation and translation, respectively. If the relative poses are time varying, a subscripts t would be used. To avoid abuses in notation of t, all translations are in **bold**.

Without loss of generality, one camera of the multiple non-overlapping camera system can be selected as the reference camera C_{ref}. The objective of the calibration approach is to estimate $\{R_{C_{ref}}^{C_i}, \mathbf{t}_{C_{ref}}^{C_i} | i = 1, 2, \ldots, n\}$. For a metric calibration, the metric scaled translation $\mathbf{t}_{C_{ref}}^{C_i}$, which denotes the physical distances in metres, should be estimated. The metric scale factor is denoted as s. An overview of the proposed method is illustrated in Fig. 2.

3.2 Relative Pose Estimation via Structure from Motion

For the non-overlapping camera systems, the lack of common FoV make it difficult to find pixel correspondences between two views, resulting in difficulty of

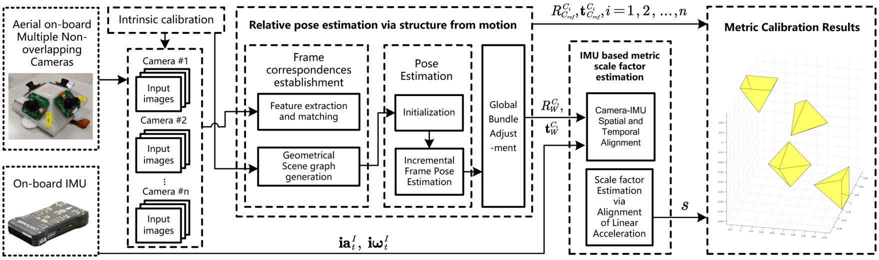

Aerial on-board Multiple Non-overlapping Cameras
On-board IMU
Intrinsic calibration
Camera #1
Input images
Camera #2
Input images
Camera #n
Input images
Relative pose estimation via structure from motion
Frame correspondences establishment
Feature extraction and matching
Geometrical Scene graph generation
Pose Estimation
Initialization
Incremental Frame Pose Estimation
Global Bundle Adjust -ment
$R_{C_{ref}}^{C_i}, \mathbf{t}_{C_{ref}}^{C_i}, i=1,2,...,n$
$R_W^{C_i}$, $\mathbf{t}_W^{C_i}$
IMU based metric scale factor estimation
Camera-IMU Spatial and Temporal Alignment
Scale factor Estimation via Alignment of Linear Acceleration
s
$\mathbf{ia}_t^I$, $\mathbf{i\omega}_t^I$
Metric Calibration Results

Fig. 2. Calibration framework of the proposed approach. The relative poses of the n cameras are estimated via SfM and global bundle adjustment. The scale ambiguity is then solved by aligning visual and inertial linear accelerations.

direct relative pose estimation. The relative rotation and translation estimation in this section is based on structure from motion, which estimates the poses of moving cameras and sparse scene points 3D locations from unordered images.

It can be assumed that the intrinsic parameters of each non-overlapping cameras are pre-calibrated and the cameras are synchronized. By jointly moving the entire system and observing a static scene, a series of images from different cameras at different times are captured, which form the input for SfM. First, image features like SIFT are extracted and matched on these images to find correspondences between images. By geometrically verifying images pairs via planar homography or epipolar geometry, a scene graph can be constructed. Images are the nodes and verified image pairs are the edges. By selecting two nodes, a series of 3D scene points can be initialized by stereo reconstruction. Other images are then incrementally added and registered to the reconstructed scene from the 2D-3D correspondences. New scene points can be added by solving the triangulation problem. Finally, a global bundle adjustment is preformed by minimizing the re-projection error of all M feature pixels among all images:

$$E_{repj} = \sum_{m=1}^{M} ||\pi(K_i, R_W^{C_{i,k}}, \mathbf{t}_W^{C_{i,k}}, \mathbf{X}_j^W) - \mathbf{x}_m||^2, \tag{1}$$

where $\pi(K_i, R_W^{C_{i,k}}, \mathbf{t}_W^{C_{i,k}}, \mathbf{X}_j^W)$ is the function to project 3D points $\mathbf{X}_j^W$ to camera C_i at time k. x_m is the actual pixel location that corresponding to the projection. After the optimization, the camera poses of the input images can be estimated.

Based on SfM, the poses of multiple cameras (n in total) in the system at different time (t in total) can be estimated with respect to the world coordinate system W, which are denoted by $\{R_W^{C_{i,k}}, \mathbf{t}_W^{C_{i,k}} | i = 1, 2, \ldots, n, k = 1, 2, \ldots, t\}$. Thus, the relative pose between camera C_i and C_{ref} at time k, $R_{C_{ref,k}}^{C_{i,k}}, \mathbf{t}_{C_{ref,k}}^{C_{i,k}}$, can be calculated by

$$R_{C_{ref,k}}^{C_{i,k}} = R_W^{C_{ref,k}} \left(R_W^{C_{i,k}} \right)^{-1}, \tag{2}$$

$$\mathbf{t}_{C_{ref,k}}^{C_{i,k}} = \mathbf{t}_W^{C_{ref,k}} - R_{C_{ref,k}}^{C_{i,k}} \mathbf{t}_W^{C_{i,k}}. \tag{3}$$

Because that cameras in the non-overlapping system are rigidly connected, it can be assumed that $\{R_{C_{ref,k}}^{C_{i,k}}, \mathbf{t}_{C_{ref,k}}^{C_{i,k}} | k = 1, 2, \ldots, t\}$ at different time are constant. Thus, the final estimation of $R_{C_{ref}}^{C_i}, \mathbf{t}_{C_{ref}}^{C_i}$ could be obtained by averaging t estimations at different times. Similar procedure is performed to obtain $\{R_{C_{ref}}^{C_i}, \mathbf{t}_{C_{ref}}^{C_i} | i = 1, 2, \ldots, n\}$. It should be noted that there would be a scale ambiguity in the estimated relative translations, which will be solved in the next subsection.

3.3 Inertial Measurement Data Based Metric Scale Factor Estimation

Based on SfM, the relative translations $\{\mathbf{t}_{C_{ref}}^{C_i} | i = 1, 2, \ldots, n\}$ are determined up to an unknown scale. To complete the metric calibration, the metric scale factor, which corresponds to the physical distances of $\{\mathbf{t}_{C_{ref}}^{C_i} | i = 1, 2, \ldots, n\}$, should be estimated. Although in some recent approaches, physical size of the checkerboard squares [22] can be used for computing the metric scale factor, we plan to solve the scale ambiguity in a pattern-free way. Our scale estimation procedure is based on Mustaniemi et al.'s approach [13], which optimizes the scale factor to align the acceleration obtained from differentiated visual poses and that from the IMU.

The linear acceleration of the system can be obtained in two ways. One can subtract the gravity vector from the raw accelerometer readings to obtain the linear acceleration in I. It is denoted by $\mathbf{ia}_t^I$. In the meanwhile, since the entire system is rigid, it can be assumed that the linear acceleration of the system is the same with that of one selected camera, say camera C_i. Hence, by differentiating the time varying positions of C_i in W, i.e., $\{\mathbf{t}_{C_{i,k}}^W, | k = 1, 2, \ldots, t\}$ twice, the linear acceleration in W can be computed. Which is denoted by $\mathbf{va}_t^W$. Considering the fact that noises in camera position are amplified, Rauch-Tung-Striebel smoother [16] is used when performing the double differentiation [13]. Note that there would usually be a scale ambiguity between $\mathbf{ia}_t^I$ and $\mathbf{va}_t^W$. By aligning the two linear accelerations, the metric scale factor can be estimated. Assuming N linear accelerations are measured in total, the objective function for scale factor estimation can be defined as

$$\arg\min_s \sum_{t=1}^{N} ||sR_{W,t}^{C_i}\mathbf{va}_t^W - R_I^{C_i}\mathbf{ia}_t^I||^2, \tag{4}$$

where $R_{W,t}^{C_i}$ is the time varying rotation between W and C_i, $R_I^{C_i}$ the relative rotation between I and C_i.

The two accelerations in Eq. (4) are usually not temporally aligned. To estimated $R_I^{C_i}$ and the time offset t_d, an objective function can be defined as

$$\arg\min_{R_I^{C_i}, \mathbf{b}_\omega^{C_i}, t_d} \sum_{t=1}^{N} ||\mathbf{v}\omega_t^{C_i} - R_I^{C_i}\mathbf{i}\omega_{t-t_d}^I + \mathbf{b}_\omega^{C_i}||^2, \tag{5}$$

where $\mathbf{v}\omega$ and $\mathbf{i}\omega$ are the angular velocity obtained from visual rotation differentiation and gyroscope, respectively. Since the outputs of a real gyroscope or a real accelerometer are usually biased, the gyroscope bias $\mathbf{b}_{\omega}^{C_i}$ in C_i is also estimated. Based on the time offset, Eq. (4) can be written as

$$\arg\min_{s,\mathbf{b}_a^{C_i}} \sum_{t=1}^{N} ||s\mathbf{R}_{W,t}^{C_i}\mathbf{va}_t^W - \mathbf{R}_I^{C_i}\mathbf{ia}_{t-t_d}^I + \mathbf{b}_a^{C_i}||^2, \tag{6}$$

where $\mathbf{b}_a^{C_i}$ denotes the accelerometer bias in C_i. In practice, the optimization in Eq. (5) is minimized using alternating optimization. Equation (6) is optimized in frequency domain. A more formal description of the implementation is given in [13]. After the estimation of metric scale factor s, we amplify the relative translations of $\{\mathbf{t}_{C_{ref}}^{C_i}|i = 1, 2, \ldots, n\}$ by s, and completes the metric calibration.

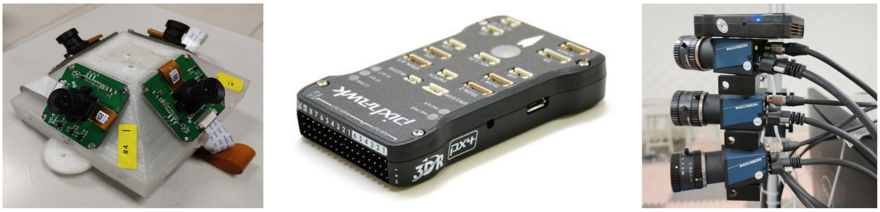
4 Experimental Results

In this section, we describe the details of the experiments to verify the accuracy of the proposed approach. First, we give the details of the two multiple non-overlapping cameras systems, together with the IMU system. Quantitative evaluations of the metric calibration results of the two systems are then given. Finally, the metric scale estimation is quantitatively evaluated via object 3D reconstruction.

4.1 Equipment

In order to verify the feasibility and accuracy of the proposed approach, an aerial non-overlapping camera system is designed and constructed. We also build a non-overlapping camera system based on industrial cameras for further verification of the calibration accuracy.

The aerial non-overlapping cameras system includes four embedded cameras, an image synchronization board, a Raspberry Pi, and a 3D-printed camera bracket. The entire system is shown in Fig. 3(a). The orientation of the four cameras are carefully designed so that there are no common FoVs between adjacent cameras. For the embedded cameras, we use OV9281 with a resolution of 1280 × 800 pixels. An Arducam four-lens camera capture board [1] is used for synchronization. For the IMU system, we directly adopt the MPU6000 chip integrated in the Pixhawk4 UAV flight controller (Fig. 3(b)) [4]. The acquisition frame rate of non-overlapping cameras in this system is designed to be 20 frame per second (FPS), while the acquisition frequency of the IMU 100 Hz. It should be noted that we do not perform the synchronization between the cameras and the IMU. The captured images and the inertial data are saved using a Raspberry Pi. The total weight of the system is around 160 g, and the power supply of the system are directly from the UAV batteries. It is convenient to adopt the system to another UAV platform.



(a) (b) (c)

Fig. 3. Equipment used in the experiment. **(a)** The aerial on-board non-overlapping camera system. **(b)** Pixhawk 4 UAV flight controller, the IMU chip provides the inertial data for both of the two system. **(c)** Industrial cameras based non-overlapping camera system.

For the industrial camera based system, we choose to use the Daheng MER-131-75GM cameras [2]. The system is shown in Fig. 3(c), in which 3 cameras are pointing at different angles for a non-overlapping FoV. The resolution of the captured image is 1280 × 1024 pixels and the frame rate is 20. A STM32ZET6 MCU is used for hardware based synchronization. We also use the IMU in Pixhawk UAV flight controller in this system, with an acquisition frequency 100 Hz. The captured images and inertial data are uploaded to a PC station for further calibration.

4.2 Metric Calibration of the Aerial On-Board Non-overlapping Camera System

In this experiment, our goal is to quantitatively evaluate the accuracy of the proposed metric calibration method. We apply the proposed approach to the aerial on-board non-overlapping camera system mentioned in Sec. 4.1. Since there are 4 cameras in total in the system, we choose one of them as the reference camera. The relative rotations and translations of all the other 3 cameras are estimated w.r.t the reference camera. We name the three camera as camera #1 to #3.

To quantitatively evaluate the accuracy of the calibration, we also perform a multiple camera based approach to obtain the ground truth of the relative rotation and metric translations. Figure 4(a) gives the details when obtaining the ground truth. By using a global camera C_{global}, which shares common FoV with both camera C_1 and C_{ref}, two checkerboards can be used to obtain the relative poses between C_1 and C_{global}, and that between C_{ref} and C_{global}. Then the relative pose between C_1 and C_{ref} can be obtained via coordinate transforms and we perform a bundle adjustment to optimize the relative poses. By utilizing multiple checkerboards, the ground truth of the relative poses between all cameras and the reference camera can be obtained.

Table 1 shows the quantitative evaluation of the metric calibration results for aerial on-board non-overlapping camera system. For each row, the roll angle, yaw angle and pitch angle of the relative orientations between camera $\{C_i | i = 1, 2, 3\}$

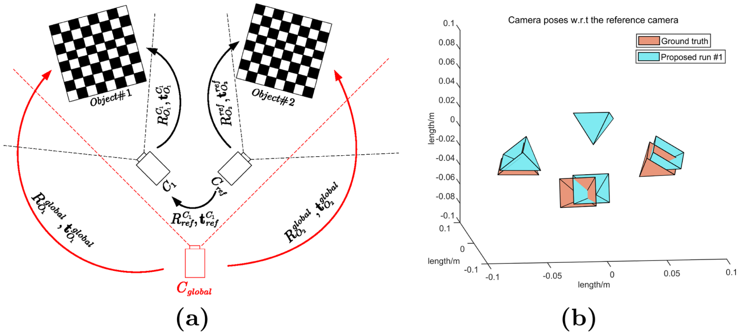

Object#1
Object#2
$R_{O_1}^{C_1}, \mathbf{t}_{O_1}^{C_1}$
$R_{O_2}^{ref}, \mathbf{t}_{O_2}^{ref}$
C_1
C_{ref}
$R_{ref}^{C_1}, \mathbf{t}_{ref}^{C_1}$
$R_{O_1}^{global}, \mathbf{t}_{O_1}^{global}$
$R_{O_2}^{global}, \mathbf{t}_{O_2}^{global}$
C_{global}
(a)
Camera poses w.r.t the reference camera
Ground truth
Proposed run #1
length/m
length/m
length/m
0.1
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
-0.1
0.1
0
-0.1
-0.1
-0.05
0
0.05
0.1
(b)

Fig. 4. Calibration results for aerial on-board non-overlapping camera system. **(a)** procedure for obtaining the ground truth **(b)** metric calibration results in run #1. (Color figure online)

and the reference camera C_{ref} is given in degree. The relative translations is given along X, Y and Z axes in metres. We run 3 different calibrations for the same system. The calibration results, together with the ground truth values are given. It can be seen that the average rotational error is less that **0.2**° and the average translational error is less than **0.015** m. Figure 4(b) illustrates the comparison between ground truth relative poses (in red color) and the estimated ones in run #1 (in blue color). Please note that in both Table 1 and Fig. 4(b), the errors in scale of the estimated translations and the ground truth are small, which shows the accuracy of the metric scale estimation.

Table 1. Estimated relative orientations and positions of different cameras w.r.t the reference camera in the **aerial on-board** non-overlapping camera system.

Approaches	camera	roll (°)	yaw (°)	pitch (°)	$\mathbf{t_x}$ (m)	$\mathbf{t_y}$ (m)	$\mathbf{t_z}$ (m)
Ground	#1	44.6052	44.9146	−1.4001	0.0283	−0.0425	−0.0364
truth	#2	90.1890	−1.4332	88.6350	0.0532	−0.0027	−0.0672
values	#3	54.7661	29.1199	124.2423	0.0217	0.0411	−0.0389
Proposed	#1	44.5014	44.9908	−1.3208	0.0342	−0.0402	−0.0332
run #1	#2	90.2865	−1.4905	88.8299	0.0535	0.0085	−0.0640
	#3	54.8420	29.0332	124.6838	0.0203	0.0462	−0.0325
Proposed	#1	44.6957	45.0860	−1.3637	0.0298	0.0337	−0.0415
run #2	#2	90.4184	−1.416	88.7724	0.0566	0.0087	−0.0643
	#3	55.1482	28.8704	124.6518	0.0197	0.0445	−0.0316
Proposed	#1	44.5544	44.8269	−1.2525	0.0231	−0.0483	−0.0306
run #3	#2	90.1434	−1.3628	88.7618	0.0531	−0.0025	−0.0654
	#3	54.9328	29.1326	124.2975	0.0181	0.0353	−0.032
Average error		0.1380	0.0921	0.1810	0.0026	0.0135	0.0047

4.3 Metric Calibration of an Industrial Non-overlapping Camera System

To further evaluate the performances of the proposed method on different camera systems, we apply the propose approach to the industrial cameras based non-overlapping camera system mentioned in Sect. 4.1. Multiple checkerboards based approach that is similar with that in Sect. 4.2 is conducted for obtaining the ground truth. In this experiment, we run the calibration for 2 different times. Table 2 gives the quantitative evaluations of the metric calibration results. It can be seen that the average rotational error is less than 0.05°, and the average translational error is less than 0.005 metres. From the comparison between Table 1 and Table 2, it can be seen that cameras with a higher resolution (1280 × 1024 pixels in Table 2 v.s. 1280 × 800 pixels in Table 1) would have a calibration result with less errors. Besides, the proposed method can be applied to different non-overlapping camera systems with a reasonable accuracy.

Table 2. Estimated relative orientations and positions of different cameras w.r.t the reference camera in the **industrial** non-overlapping camera system.

Approaches	camera	roll (°)	yaw (°)	pitch (°)	$\mathbf{t_x}$ (m)	$\mathbf{t_y}$ (m)	$\mathbf{t_z}$ (m)
Ground truth values	#1	2.1209	14.3206	0.8938	0.0147	−0.0581	−0.0148
	#2	3.0223	35.1218	2.1217	0.0252	−0.1075	−0.0252
Proposed run #1	#1	2.0896	14.2859	0.8889	0.0174	−0.0557	−0.0102
	#2	2.9402	35.1025	2.1134	0.0237	−0.1047	−0.0237
Proposed run #2	#1	2.1076	14.2665	0.8773	0.0198	−0.0617	−0.0198
	#2	2.9507	35.0922	2.0832	0.0317	−0.1161	−0.0215
Average error		0.0496	0.0344	0.0170	0.0040	0.0043	0.0037

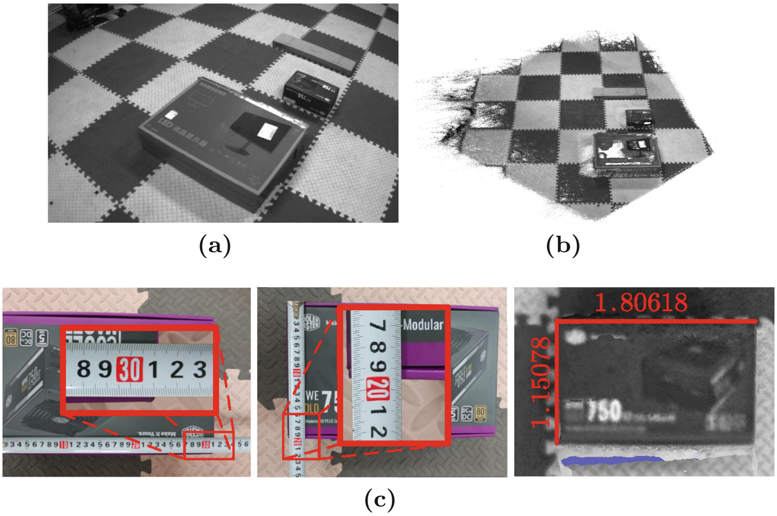
4.4 Experiments of Applications for Object Metric 3D Reconstruction

Because the metric relative translations are estimated, one can obtain a metric reconstruction of scene objects purely based on images from the systems. No external metric sensors, such as the depth sensor or the inertial sensor, is needed in the metric reconstruction.

In this experiment, we give a quantitatively evaluation of the metric scale in this application. We place 3 different boxes in a static scene (Fig. 5(a)). By moving the on-board non-overlapping cameras and observing several static scene objects, a dense 3D reconstruction can be obtained via SfM [17] (Fig. 5(b)). The length of a object in the scene can be obtained using Meshlab [3] (Fig. 5(c)). Based on the metric scale estimated from the proposed method, the 3D reconstruction can be scaled to the physical scale. We compare the estimated object length with the ground truth length obtained from a ruler (Fig. 5(c)). Table 3 gives the quantitative evaluation of the estimated object lengths or widths in metres. It can be seen that the errors are all less than 2%, which shows the accuracy of the estimated metric scale factor in the proposed approach.

Table 3. Quantitative evaluation of the metric 3D reconstruction on 3 different scene objects. The scale factor is 0.1713 in this experiment.

Object		Ruler (m)	Reconstructed	Estimated (m)	Error
#1	Length	0.7050	4.1589	0.7124	1.06%
	Width	0.5190	3.0546	0.5233	0.81%
#2	Length	0.3110	1.8062	0.3094	0.35%
	Width	0.1950	1.1508	0.1971	1.34%
#3	Length	0.6720	3.9714	0.6803	1.20%
	Width	0.1650	0.9496	0.1627	1.51%



(a)

(b)

(c)

Fig. 5. Results of applications for object metric 3D reconstruction. **(a)** A static scene with 3 boxes. **(b)** Reconstructed point cloud. **(c)** Length and width of object #2 obtained from both rulers and the point clouds.

5 Conclusions

In this work, a novel metric calibration method for aerial on-board multiple non-overlapping cameras is proposed. The 3D geometry and consistence of a static scene is utilized to establish the pixel correspondences between non-overlapping cameras. To overcome the scale problem in SfM based pose estimation, the inertial measurement data is used for estimate the metric scale. Real experiment of both aerial on-board and industrial camera based non-overlapping camera systems shows the accuracy of our approach. No calibration object is used during the calibration procedure. In the future work, we would like to consider the application of aerial non-overlapping cameras for remote sensing, metric scene reconstruction, and active object detections.

References

1. ArduCam. https://www.arducam.com
2. Daheng Imaging cameras. https://www.daheng-imaging.com/
3. Meshlab. https://www.meshlab.net/
4. Pixhawk flight controller. https://pixhawk.org/products/
5. Anderson, R., et al.: Jump: virtual reality video. ACM Trans. Graph. (TOG) **35**(6), 1–13 (2016)
6. Caspi, Y., Irani, M.: Aligning non-overlapping sequences. Int. J. Comput. Vis. (IJCV) **48**(1), 39–51 (2002)
7. Dong, S., Shao, X., Kang, X., Yang, F., He, X.: Extrinsic calibration of a non-overlapping camera network based on close-range photogrammetry. Appl. Opt. **55**(23), 6363–6370 (2016)
8. Gong, Z., Liu, Z., Zhang, G.: Flexible global calibration of multiple cameras with nonoverlapping fields of view using circular targets. Appl. Opt. **56**(11), 3122–3131 (2017)
9. Kassebaum, J., Bulusu, N., Feng, W.C.: 3-D target-based distributed smart camera network localization. IEEE Trans. Image Process. (TIP) **19**(10), 2530–2539 (2010)
10. Kumar, R.K., Ilie, A., Frahm, J.M., Pollefeys, M.: Simple calibration of non-overlapping cameras with a mirror. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–7. IEEE (2008)
11. Li, B., Heng, L., Koser, K., Pollefeys, M.: A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1301–1307. IEEE (2013)
12. Liu, Z., Zhang, G., Wei, Z., Sun, J.: A global calibration method for multiple vision sensors based on multiple targets. Meas. Sci. Technol. **22**(12), 125102 (2011)
13. Mustaniemi, J., Kannala, J., Särkkä, S., Matas, J., Heikkilä, J.: Inertial-based scale estimation for structure from motion on mobile devices. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4394–4401. IEEE (2017)
14. Pagel, F.: Extrinsic self-calibration of multiple cameras with non-overlapping views in vehicles. In: Video Surveillance and Transportation Imaging Applications 2014, vol. 9026, p. 902606. International Society for Optics and Photonics (2014)
15. Pagel, F., Willersinn, D.: Motion-based online calibration for non-overlapping camera views. In: 13th International IEEE Conference on Intelligent Transportation Systems, pp. 843–848. IEEE (2010)
16. Rauch, H.E., Tung, F., Striebel, C.T.: Maximum likelihood estimates of linear dynamic systems. AIAA J. **3**(8), 1445–1450 (1965)
17. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4104–4113 (2016)
18. Strauß, T., Ziegler, J., Beck, J.: Calibrating multiple cameras with non-overlapping views using coded checkerboard targets. In: 17th International IEEE Conference on Intelligent Transportation Systems, pp. 2623–2628. IEEE (2014)
19. Sturm, P., Bonfort, T.: How to compute the pose of an object without a direct view? In: Narayanan, P.J., Nayar, S.K., Shum, H.Y. (eds.) ACCV 2006. LNCS, vol. 3852, pp. 21–31. Springer, Heidelberg (2006). https://doi.org/10.1007/11612704_3
20. Svoboda, T., Martinec, D., Pajdla, T.: A convenient multicamera self-calibration for virtual environments. Presence Teleoperators Virtual Environ. **14**(4), 407–422 (2005)

21. Xing, Z., Yu, J., Ma, Y.: A new calibration technique for multi-camera systems of limited overlapping field-of-views. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5892–5899. IEEE (2017)
22. Yin, L., Wang, X., Ni, Y., Zhou, K., Zhang, J.: Extrinsic parameters calibration method of cameras with non-overlapping fields of view in airborne remote sensing. Remote Sens. **10**(8), 1298 (2018)
23. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) **22**(11), 1330–1334 (2000)
24. Zhu, C., Zhou, Z., Xing, Z., Dong, Y., Ma, Y., Yu, J.: Robust plane-based calibration of multiple non-overlapping cameras. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 658–666. IEEE (2016)