Minimal Solutions for the Rotational Alignment of IMU-Camera Systems using Homography Constraints☆

Banglei Guan[a,b,*], Qifeng Yu[a,b], Friedrich Fraundorfer[c]

[a] *College of Aerospace Science and Engineering, National University of Defense Technology*
[b] *Hunan Provincial Key Laboratory of Image Measurement and Vision Navigation*
[c] *Institute for Computer Graphics and Vision, Graz University of Technology*

Abstract

In this paper, we explore the different minimal case solutions to the rotational alignment of IMU-camera systems using homography constraints. The assumption that a ground plane is visible in the images can easily be created in many situations. This calibration process is relevant to many smart devices equipped with a camera and an inertial measurement unit (IMU), like micro aerial vehicles (MAVs), smartphones and tablets, and it is a fundamental step for vision and IMU data fusion. Our solutions are novel as they compute the rotational alignment of IMU-camera systems by utilizing a first-order rotation approximation and by solving a polynomial equation system derived from homography constraints. These solutions depend on the calibration case with respect to camera motion (general motion case or pure rotation case) and camera parameters (calibrated camera or partially uncalibrated camera). We then demonstrate that the number of matched points in an image pair can vary from 1.5 to 3. This enables us to calibrate using only one relative movement and provide the exact algebraic solution to the problem. The novel minimal case solutions are useful to reduce the computation time and increase the calibration robustness when using Random Sample Consensus (RANSAC) on the point correspondences between two images. Furthermore, a non-linear parameter optimization over all image pairs is performed. In contrast to the previous calibration methods, our solutions do not require any special hardware,

☆The research has been performed during a research visit under the supervision of Friedrich Fraundorfer at Graz University of Technology.

*Corresponding author
Email address: banglei.guan@hotmail.com (Banglei Guan)

and no problems are experienced with one image pair without special motion. Finally, by evaluating our algorithm on both synthetic and real scene data including data obtained from robots, smartphones and MAVs, we demonstrate that our methods are both efficient and numerically stable for the rotational alignment of IMU-camera systems.
Keywords: IMU-camera calibration, Rotational alignment, Minimal Solution, Homography constraint, Algebraic solution, Pure rotation

1. Introduction

With the omnipresence of smart devices, the fusion of vision and IMU data play an important role in a wide variety of applications such as simultaneous localization and mapping (SLAM) [1] and structure from motion (SfM) [2, 3]. In order to perform data fusion, IMU-camera calibration must be performed in advance to determine the transformation between the IMU coordinate system and the camera coordinate system, which consists of a rotational component and a translational component.

For many applications only the rotational alignment is of importance, and the translational component between the IMU and the camera coordinate systems does not need to be calibrated. Example applications are up-righting photos on a smart phone or special instances of visual-inertial ego-motion estimation [3, 4]. However, the accuracy that can be achieved with these applications highly depends on the axis alignment between the IMU and the camera coordinate system. Therefore, this paper focuses on the rotational alignment of IMU-camera systems.

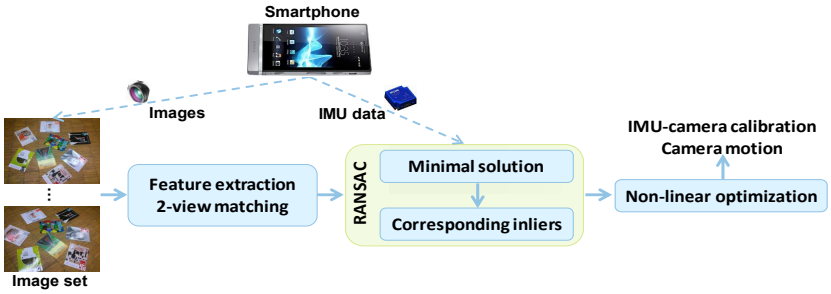


Figure 1: Overview of the proposed IMU-camera calibration methods. Our methods not only can be used to calibrate the rotational component between the IMU and the camera using a single image pair, but also can be used to achieve robust calibration results using RANSAC on multiple image pairs through exhaustive pairwise matching.

IMU-camera calibration can be regarded as hand-eye calibration regarding the IMU as the hand [5, 6, 7, 8, 9], which has been widely considered for robotic or automotive applications. Most of these methods compute the hand-eye calibration from rigid transformation matrices of subsequent time steps. In our work, we propose to compute the IMU-camera calibration directly from feature matches and also propose a robust estimator by utilizing the RANSAC [10] to cope with outliers in the data. For such a RANSAC scheme, a minimal case solution is of the utmost importance, because the number of random samples that must be taken to find one outlier free sample depends exponentially on the number of parameters to instantiate one hypothesis. The goal of this paper is to describe a technique allowing the rotational alignment of IMU-camera systems to be performed robustly and accurately. Figure 1 illustrates the proposed IMU-camera calibration methods. We derive different minimal solutions depending on the calibration case:

- If the motion of the calibrated camera is general motion including rotation and translation, we develop a minimal solution using 3 point correspondences. The solution is novel as it computes the camera motion and the IMU-camera calibration simultaneously.

- If the motion of the calibrated camera is a pure rotation or can be approximated effectively as a pure rotation, we will see that only 1.5 point correspondences are

required to calibrate the rotational component between the camera and the IMU.

- If there is a partially uncalibrated camera, whereby the intrinsic parameters except the focal length are known, and the motion of the camera is a pure rotation, we use 2 point correspondences to retrieve the focal length of the camera and IMU-camera calibration.

Our contributions can be summarized in the following way:

- We derive the minimal case solutions for the rotational alignment of IMU-camera systems using homography constraints. By applying a first-order approximation of the rotation (when the three installation angles between the IMU and the camera are approximately known), a practically usable implementation could be found. These methods are efficient within a RANSAC scheme, and they can also be effectively used to perform IMU-camera calibration on devices with limited computational power (*e.g.* smartphones and tablets).

- The proposed methods remove the requirement for the prior knowledge of the camera poses. We directly minimize the image transfer residuals based on homography constraints, rather than conduct an algebraic minimization of transformation matrices between the IMU and the camera. The objective function based on the image measurements is a geometrically more meaningful criterion.

- Our solutions are novel as they allow us to compute the camera motion and the IMU-camera calibration simultaneously without using a known calibration device or any special hardware.

The proposed methods are evaluated on synthetic and real data sets. We test the algorithms under different levels of rotation magnitude and image noise. The synthetic results show that our solutions do not show a significant loss in accuracy when operating under the assumption of first-order rotation approximation. We conduct a detailed analysis of real data sets, including a robotic data set, a MAV data set and a common smartphone data set, and compare the results with these from state-of-the-art methods. In particular, we demonstrate the use of the proposed methods under the challenging

condition of only using a small number of images for calibration. Further, we evaluate the accuracy obtained using different calibration methods with the ground truth. The calibration results confirm the validity and robustness of the proposed IMU-camera calibration methods.

The remainder of the paper is structured as follows. First, we review related work in Section 2. In Section 3, we establish basics and notations for IMU-camera calibration methods using homography constraints. In Section 4, we derive the different minimal case solutions using the Gröbner basis technique or analytical method according to the calibration case and describe the non-linear parameter optimization over all image pairs. In Section 5, we validate the methods experimentally using both synthetic and real scene data. Finally, concluding remarks are given in Section 6.

2. Related work

The IMU-camera calibration problem and the related hand-eye calibration problem have already been addressed by various authors in many papers. A class of approaches to this problem use a filter-based approach to estimate the calibration information as part of visual-inertial sensor fusion [11, 12, 13]. These approaches use inertial measurements directly and consider the correlations between the IMU measurements. A high camera frame rate is required because of the large number of DOFs in those approaches.

In addition, some methods address the problem of rolling shutters in camera sensors due to high frame rate [14, 15, 16, 17]. Rolling shutter constraints are important for calibrating from video sequences, where the camera is moving during acquisition. But in our case, we are taking images but not recording the video. There is not fast relative motion between the scene and the camera. Even when we use a rolling shutter camera, we will acquire still images. In addition, we require a static scene, so a rolling shutter camera will not produce artifacts with a non-moving camera and a static scene. The rolling shutter effect is not necessary to taken into account in our paper. Moreover, some methods require knowledge about the properties of the scene, *e.g.* known calibration targets [12, 18, 19]. In contrast, a method for IMU-camera calibration without the

use of a known calibration device or any special hardware is presented in this paper, which is useful in many situations where calibration device or special hardware will not be allowed or provided.

Typically, common IMUs output the complete rotation information with respect to the IMU reference coordinate system. Hence, in the following review, we focus on the different approaches which directly use rotation information from the IMU to solve the IMU-camera calibration problem. Hand-eye calibration has been studied by many researchers in the past. The standard formulation of the hand-eye calibration problem leads to a solution to the well-known equation $\mathbf{AX} = \mathbf{XB}$, where $\mathbf{A}$ and $\mathbf{B}$ are the known relative rigid motions of the camera and the IMU, respectively. It has been shown that the transformation between the IMU coordinate system and the camera coordinate system $\mathbf{X}$ can be determined with at least two motions along non-parallel rotation axes [9]. The existing methods can be divided into three groups. The first group of methods solves the rotational and translational components separately [6, 8, 9] or only solves the rotational component [20]. The second group of methods solves the rotational and translational components simultaneously [5, 6, 7, 21]. Kukelova *et al.* [7] presented the minimal problem of hand-eye calibration for the situations, whereby the translational components of the hand can be measured but rotational components are not known. The transformation $\mathbf{X}$ is solved by the minimal number of two relative movements, and the solution can be refined afterward by applying the optimization method of Zhuang and Shiu [21]. However, both groups of methods require the prior knowledge of the camera poses, which are recovered by a calibration pattern or a SfM approach.

Recently, another group of methods has been described that use image measurements directly and do not require prior knowledge of the camera poses. Ruland *et al.*[22] and Heller *et al.*[23, 24] solved for the rotation and translation simultaneously by minimizing the residuals in image space. The above-mentioned methods employ the branch-and-bound algorithm to obtain a globally-optimal estimate with respect to L_∞-norm minimization. As these methods have not adopted any procedures to cope with outliers, their accuracy is highly influenced by feature mismatches. For an aircraft equipped with a camera and a GPS-corrected inertial navigation system, Ben-

der *et al.*[25] performed an in-flight calibration of camera parameters and boresight with a graph optimization framework. Moreover, for smart devices like smartphones and tablets, the IMU alone cannot provide the position information as the actuator of robots can. The approaches which solve simultaneously the rotational and translational components cannot be used.

Moreover, there is a special IMU-camera calibration situation where the motion of the camera is assumed to be pure rotation. Seo *et al.*[26] assumed all the translations to be zero and solved the rotational component between the IMU and the camera using image correspondences. Hwangbo *et al.*[27] also presented a calibration method based on homography transformation of image correspondences assuming pure rotation. Karpenko *et al.*[28] calibrated the camera and gyroscope system from a single input video, which was obtained by quickly shaking the camera while pointing it at a far-away object. Pure rotation case has practical relevance. By rotating the camera outside, where everything is far away, the parallax-shift of most objects is hardly noticeable. Such data is close enough to a pure rotation case such that an algorithm for a pure rotation case can be applied to it. In the pure rotation case, it also has already been established that it is possible to recover the focal length of the camera [29].

Mathematically, given the relative rotations as measured in two coordinate frames, the relative rotation between two coordinate frames can be found using the Procrustes method [30]. As the relative camera rotations in this case are computed from image features, the accuracy of the relative rotations depend on the image features. In [31], error propagation was used to analyze the dependency on the quality of the image features, and it has been stated that the method which directly minimizes the image transfer residuals leads to better results than the Procrustes method.

In this work, rather than computing essential matrices to extract the relative motions, the relative motions are extracted from homographies computed between image pairs. The estimation of a homography needs fewer point correspondences than the estimation of the essential matrix, which is beneficial for use in a RANSAC loop. The assumption that a ground plane is visible in the images lowers again the number of necessary point correspondences, while this condition can easily be created in many situations.

Ventura *et al.*[32] propose a minimal solution for estimating the motion of a multi-camera rig by using a first-order approximation to relative pose. In many practical cases, the approximate installation relationship between the IMU and the camera is known (*e.g.* hand-measured or extracted from device layouts). Therefore, we can safely use a first-order approximation of the rotation matrix, which simplifies the IMU-camera calibration problem and allows us to find minimal case solutions.

3. Basics and notations

With known intrinsic camera parameters, a general homography relation between two different views is represented as follows [33]:

$$\lambda \mathbf{x}_j = \mathbf{H}\mathbf{x}_i = (\mathbf{R} - \frac{1}{d}\mathbf{t}\mathbf{N}^T)\mathbf{x}_i, \tag{1}$$

where $\mathbf{x}_i = [x_i, y_i, 1]^T$ and $\mathbf{x}_j = [x_j, y_j, 1]^T$ are the normalized homogeneous image coordinates of the points in views i and j, and λ is a scale factor. $\mathbf{H}$ is the homography matrix, $\mathbf{R}$ and $\mathbf{t}$ are the rotation and the translation from views i to j, respectively, and d is the distance between the view i frame and the 3D plane. $\mathbf{N}$ is the unit normal vector of the 3D plane with respect to the view i frame.

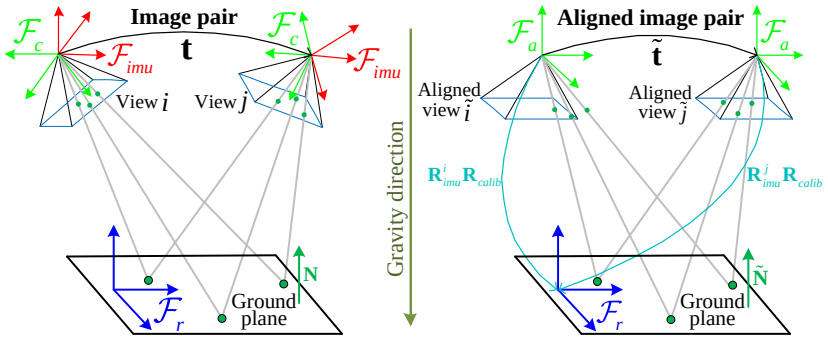


Figure 2: The illustration of the camera coordinate system $\mathcal{F}_c$, IMU coordinate system $\mathcal{F}_{imu}$, IMU reference coordinate system $\mathcal{F}_r$ and aligned camera coordinate system $\mathcal{F}_a$. We show both a general image pair (left) and aligned image pair (right).

As shown in Figure 2, the rotational alignment difference between $\mathcal{F}_c$ and $\mathcal{F}_{imu}$ is

expressed with $\mathbf{R}_{calib}$, while the orientation estimations from $\mathcal{F}_{imu}$ to $\mathcal{F}_r$ are given by $\mathbf{R}_{imu}$. The rotations of views i and j can be expressed as $\mathbf{R}^i_{imu}\mathbf{R}_{calib}$, $\mathbf{R}^j_{imu}\mathbf{R}_{calib}$ in $\mathcal{F}_r$, respectively. We can align image features from $\mathcal{F}_c$ to $\mathcal{F}_a$, which coincides with the coordinate axes of $\mathcal{F}_r$. Then the relationship between two aligned views $\tilde{i}$ and $\tilde{j}$ only has a translation component $\tilde{\mathbf{t}}$ left. The unit normal vector of the ground plane with respect to the aligned view $\tilde{i}$ is expressed as $\tilde{\mathbf{N}} = [0, 0, 1]^T$. The homography relation between views i and j can be re-expressed as:

$$\lambda \mathbf{x}_j = (\mathbf{R}^T_{calib}(\mathbf{R}^j_{imu})^T\mathbf{R}^i_{imu}\mathbf{R}_{calib} - \frac{1}{d}\mathbf{R}^T_{calib}(\mathbf{R}^j_{imu})^T\tilde{\mathbf{t}}\tilde{\mathbf{N}}^T\mathbf{R}^i_{imu}\mathbf{R}_{calib})\mathbf{x}_i. \quad (2)$$

Note that in $\mathbf{t} = \mathbf{R}^T_{calib}(\mathbf{R}^j_{imu})^T\tilde{\mathbf{t}}$, the camera-plane distance d is set to 1 and absorbed by $\mathbf{t}$ [3]. By this the homography between views i and j can be rewritten as:

$$\mathbf{H} = \mathbf{R}^T_{calib}(\mathbf{R}^j_{imu})^T\mathbf{R}^i_{imu}\mathbf{R}_{calib} - \mathbf{t}\tilde{\mathbf{N}}^T\mathbf{R}^i_{imu}\mathbf{R}_{calib}. \quad (3)$$

In order to further eliminate the unknown scale factor λ, we multiply both sides of Eq. 1 by the skew-symmetric matrix $[\mathbf{x}_j]_\times$, which yields the equation:

$$[\mathbf{x}_j]_\times \mathbf{H}\mathbf{x}_i = \mathbf{0}. \quad (4)$$

Eq. 4 has three rows and only imposes two independent constraints on $\mathbf{H}$. Moreover, we exploit the fact that image correspondences are still related by homography when the motion of the camera between two views is a pure rotation or can be effectively approximated as a pure rotation. In this way, we also consider the special IMU-camera calibration case that the translation $\mathbf{t}$ from views i to j is zero.

4. IMU-camera calibration using homography constraints

This section describes the proposed algorithms for the rotational alignment of IMU-camera systems using a homography formulation. In particular we describe the derivation of polynomial equation systems to be used to compute the unknown rotational alignment parameters. We describe how these polynomial equation systems can be solved by making use of a Gröbner basis solver or in a specific case by making use of the 3Q3 method.

In the following subsections, we give the derivation of the 3pt algorithm for the calibrated camera for a general motion case. Then we give the derivation of the 1.5pt algorithm for the calibrated camera for a pure rotation case. Finally, we give the derivation of the 2pt algorithm for the partially uncalibrated camera (unknown focal length) for the pure rotation case.

4.1. 3pt calibration method for the general motion case

By parametrizing $\mathbf{R}_{calib}$ by three rotations (x, y, z) and substituting it into Eq. 3, we attain polynomial equations with 9 unknowns, *i.e.* 6 rotation parameters $\mathbf{r} = [\cos(x), \sin(x), \cos(y), \sin(y), \cos(z), \sin(z)]^T$, and 3 translation parameters for $\mathbf{t} = [t_x, t_y, t_z]^T$. Each point correspondence gives 2 linearly independent equations based on Eq. 4. The equations from 3 point correspondences give a total of 6 polynomial equations:

$$f_m(\mathbf{r}, t_x, t_y, t_z) = 0, \;\; m = 1, 2...6. \tag{5}$$

The three additional trigonometric constraints in rotation parameters $\mathbf{r}$ can be utilized:

$$\begin{aligned} \cos^2(x) + \sin^2(x) &= 1, \\ \cos^2(y) + \sin^2(y) &= 1, \\ \cos^2(z) + \sin^2(z) &= 1. \end{aligned} \tag{6}$$

Combining Eqs. 5 and 6, we attain 9 polynomial equations in 9 unknowns. A suitable way to find an algebraic solution to such a polynomial equation system is to use the Gröbner basis technique [34]. We use the automatic Gröbner basis solver described by Kukelova *et al.*[35]. Evaluating the Gröbner basis, we find that the polynomial equation system has a maximum polynomial degree of 6 and up to 48 solutions. The produced Matlab-code indicates the number of operations necessary, which involves equations that need 18018 lines to print them out, which leads to an extremely long runtime for the solver. We also experienced numerical stability issues with this derivation. As this solver should be used within a RANSAC loop, it is important to find a faster solver, especially to perform IMU-camera calibration on smart devices with limited computational power.

Our key observation is that in smart devices such as smartphones and tablets, the approximate installation relationship between the IMU and the camera which is defined as $\mathbf{R}_A$ is known from hand measurements or obtained from device layouts and is usually approximated with 0°, $\pm 90^\circ$ or 180°. We can safely approximate the rotation matrix to the first-order, which simplifies the polynomial equation system. First, we rotate the image features in views i and j using the approximate installation relationship $\mathbf{R}_A$:

$$\hat{\mathbf{x}}_i = \mathbf{R}_A \mathbf{x}_i, \quad \hat{\mathbf{x}}_j = \mathbf{R}_A \mathbf{x}_j. \tag{7}$$

The remaining rotation between the IMU coordinate system and the rotated camera coordinate system is small. This allows us to replace the remaining rotation matrix $\hat{\mathbf{R}}_{calib}$ by its first-order expansion:

$$\hat{\mathbf{R}}_{calib} = \mathbf{I}_{3\times 3} + [\hat{\mathbf{r}}]_\times, \tag{8}$$

where $\hat{\mathbf{r}} = [\hat{r}_x, \hat{r}_y, \hat{r}_z]^T$ is a three-dimensional vector. The corresponding exact rotation matrix can be retrieved by projecting the matrix to the closest rotation matrix. Like Eq. 3 and Eq. 4, we attain the new homography equation and homography constraints for the rotated image features:

$$\hat{\mathbf{H}} = \hat{\mathbf{R}}_{calib}^T (\mathbf{R}_{imu}^j)^T \mathbf{R}_{imu}^i \hat{\mathbf{R}}_{calib} - \hat{\mathbf{t}} \tilde{\mathbf{N}}^T \mathbf{R}_{imu}^i \hat{\mathbf{R}}_{calib}, \tag{9}$$

$$[\hat{\mathbf{x}}_j]_\times \hat{\mathbf{H}} \hat{\mathbf{x}}_i = \mathbf{0}. \tag{10}$$

The unknowns we are seeking for are the calibration parameters $\hat{\mathbf{r}} = [\hat{r}_x, \hat{r}_y, \hat{r}_z]^T$ and the translation $\hat{\mathbf{t}} = [\hat{t}_x, \hat{t}_y, \hat{t}_z]^T$ from the rotated views i to j. Based on Eq. 10, the equations from 3 point correspondences give a total of 6 polynomial equations:

$$f_w(\hat{r}_x, \hat{r}_y, \hat{r}_z, \hat{t}_x, \hat{t}_y, \hat{t}_z) = 0, \;\; w = 1, 2...6. \tag{11}$$

The automatic Gröbner basis solver [35] shows that this polynomial equation system has a maximum polynomial degree of 2 and at most 24 solutions. This equation system only needs 766 lines to print out. We use each solution to compose the homography for the rotated image features with Eq. 9 and choose the solution which has the

maximum number of inliers in the RANSAC loop. From this robust estimation procedure, we obtain $\hat{\mathbf{R}}_{calib}$ and $\hat{\mathbf{t}}$ for each image pair. We finally calculate the rotational component $\mathbf{R}_{calib}$ between the IMU and the camera with:

$$\mathbf{R}_{calib} = \hat{\mathbf{R}}_{calib}\mathbf{R}_A. \tag{12}$$

At the same time, the camera motion is recovered as well. The relative motion between views i and j in Eq. 1 is calculated by:

$$\mathbf{R} = \mathbf{R}_{calib}^T(\mathbf{R}_{imu}^j)^T\mathbf{R}_{imu}^i\mathbf{R}_{calib}, \tag{13}$$

$$\mathbf{t} = \mathbf{R}_A^T\hat{\mathbf{t}}. \tag{14}$$

4.2. 1.5pt calibration method for the pure rotation case

By again using the first-order approximation of the rotation, we propose two methods to perform IMU-camera calibration for the pure rotation case with the calibrated camera, specifically, the Gröbner basis method and the proposed analytical solver called the 3Q3 method.

4.2.1. Gröbner basis method

Assuming that $\hat{\mathbf{t}}$ is $[0, 0, 0]^T$ in Eq. 9, the homography matrix $\hat{\mathbf{H}}$ with pure rotation case is given by:

$$\hat{\mathbf{H}} = \hat{\mathbf{R}}_{calib}^T(\mathbf{R}_{imu}^j)^T\mathbf{R}_{imu}^i\hat{\mathbf{R}}_{calib}, \tag{15}$$

The unknowns we are seeking for are the calibration parameters $\hat{\mathbf{r}} = [\hat{r}_x, \hat{r}_y, \hat{r}_z]^T$. According to the homography constraints in Eq. 10, the equations from 1.5 point correspondences give a total of 3 polynomial equations:

$$f_w(\hat{r}_x, \hat{r}_y, \hat{r}_z) = 0, \;\; w = 1, 2, 3. \tag{16}$$

The Gröbner basis solver [35] shows that this polynomial equation system has a maximum polynomial degree of 2 and at most 8 solutions. This equation system only needs 151 lines to print out. An interesting fact in this case is that only one of the two available equations from the second point is used. Although the RANSAC loop

requires us to sample 2 points for this method, it is now possible to run a consistency check on the second point correspondence. To identify an outlier free homography hypothesis, one remaining equation has also to be fulfilled. We choose the solution which has the maximum number of inliers in the RANSAC loop, then we finally attain the rotational component $\mathbf{R}_{calib}$ between the IMU and the camera by Eq. 12.

4.2.2. 3Q3 method

The IMU-camera calibration for the pure rotation case can be formulated as the 3Q3 problem [36], which contains three quadratic equations with three unknowns. Now, we denote the problem of solving the three quadrics with three unknowns and propose the analytical solver as a 3Q3 solver.

Now, we expand the equations Eq.10 and 15 on the unknowns $\hat{\mathbf{r}} = [\hat{r}_x, \hat{r}_y, \hat{r}_z]^T$:

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} & c_{17} & c_{18} & c_{19} & c_{110} \\ c_{21} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} & c_{27} & c_{28} & c_{29} & c_{210} \end{bmatrix} \begin{bmatrix} \hat{r}_x^2 \\ \hat{r}_y^2 \\ \hat{r}_z^2 \\ \hat{r}_x\hat{r}_y \\ \hat{r}_x\hat{r}_z \\ \hat{r}_y\hat{r}_z \\ \hat{r}_x \\ \hat{r}_y \\ \hat{r}_z \\ 1 \end{bmatrix} = \mathbf{0}, \tag{17}$$

with:

$$\Delta\mathbf{R}_{IMU} = \mathbf{R}_{IMUj}^T\mathbf{R}_{IMUi} = \begin{bmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{bmatrix} \tag{18}$$

$$
\begin{cases}
c_{11} = I_{32} + I_{22}y_j - I_{33}y_i - I_{23}y_jy_i \\
c_{12} = I_{11}y_j - I_{13}y_jx_i \\
c_{13} = -I_{11}y_i + I_{12}x_i \\
c_{14} = -I_{31} - I_{12}y_j - I_{21}y_j + I_{33}x_i + I_{13}y_jy_i + I_{23}y_jx_i \\
c_{15} = -I_{12} - I_{32}x_i + I_{13}y_i + I_{31}y_i - I_{22}y_jx_i + I_{21}y_jy_i \\
c_{16} = I_{11} - I_{13}x_i + I_{12}y_jx_i - I_{11}y_jy_i \\
c_{17} = I_{22} - I_{33} - I_{23}y_j - I_{32}y_j - I_{31}x_i - I_{23}y_i - I_{32}y_i \\
\quad - I_{21}y_jx_i - I_{22}y_jy_i + I_{33}y_jy_i \\
c_{18} = -I_{21} + I_{13}y_j + I_{31}y_j + I_{23}x_i + I_{11}y_jx_i - I_{33}y_jx_i + I_{12}y_jy_i \\
c_{19} = I_{13} + I_{11}x_i - I_{22}x_i + I_{12}y_i + I_{21}y_i + I_{32}y_jx_i - I_{31}y_jy_i \\
c_{110} = I_{33}y_j - I_{23} - I_{21}x_i - I_{22}y_i + I_{31}y_jx_i + I_{32}y_jy_i
\end{cases}
\tag{19}
$$

$$
\begin{cases}
c_{21} = -I_{22}x_j + I_{23}x_jy_i \\
c_{22} = -I_{31} + I_{33}x_i - I_{11}x_j + I_{13}x_jx_i \\
c_{23} = I_{22}x_i - I_{21}y_i \\
c_{24} = I_{32} + I_{12}x_j + I_{21}x_j - I_{33}y_i - I_{23}x_jx_i - I_{13}x_jy_i \\
c_{25} = -I_{22} + I_{23}y_i + I_{22}x_jx_i - I_{21}x_jy_i \\
c_{26} = I_{21} - I_{23}x_i - I_{32}x_i + I_{31}y_i - I_{12}x_jx_i + I_{11}x_jy_i \\
c_{27} = -I_{12} + I_{23}x_j + I_{32}x_j + I_{13}y_i + I_{21}x_jx_i + I_{22}x_jy_i - I_{33}x_jy_i \\
c_{28} = I_{11} - I_{33} - I_{13}x_i - I_{31}x_i - I_{32}y_i - I_{11}x_jx_i + I_{33}x_jx_i - I_{12}x_jy_i \\
\quad - I_{13}x_j - I_{31}x_j \\
c_{29} = I_{23} + I_{12}x_i + I_{21}x_i - I_{11}y_i + I_{22}y_i - I_{32}x_jx_i + I_{31}x_jy_i \\
c_{210} = I_{13} - I_{33}x_j + I_{11}x_i + I_{12}y_i - I_{31}x_jx_i - I_{32}x_jy_i
\end{cases}
\tag{20}
$$

The equations have a maximum polynomial degree of 2. Using 1.5 points, we can compute the three unknowns $(\hat{r}_x, \hat{r}_y, \hat{r}_z)$ based on three equations. We use the two constraint equations of the first point and the first constraint equation of the second

point. The polynomial equation system can be expressed as follows:

$$\mathbf{c}_i \begin{bmatrix} \hat{r}_x^2 & \hat{r}_y^2 & \hat{r}_z^2 & \hat{r}_x\hat{r}_y & \hat{r}_x\hat{r}_z & \hat{r}_y\hat{r}_z & \hat{r}_x & \hat{r}_y & \hat{r}_z & 1 \end{bmatrix}^T = \mathbf{0}, \tag{21}$$

where the problem coefficients are $c_{ij}, i = 1, 2, 3, j = 1, 2....., 10$. We 'hide' the unknown $\hat{r}_x$, which leaves us with two unknowns $\hat{r}_y, \hat{r}_z$, and Eq. 21 can be rewritten [36]:

$$\begin{bmatrix} s_{11}^{[2]}(\hat{r}_x) & s_{12}^{[2]}(\hat{r}_x) & s_{13}^{[3]}(\hat{r}_x) \\ s_{21}^{[2]}(\hat{r}_x) & s_{22}^{[2]}(\hat{r}_x) & s_{23}^{[3]}(\hat{r}_x) \\ s_{31}^{[3]}(\hat{r}_x) & s_{32}^{[3]}(\hat{r}_x) & s_{33}^{[4]}(\hat{r}_x) \end{bmatrix} \begin{bmatrix} \hat{r}_y \\ \hat{r}_z \\ 1 \end{bmatrix} = \mathbf{M}(\hat{r}_x) \begin{bmatrix} \hat{r}_y \\ \hat{r}_z \\ 1 \end{bmatrix} = \mathbf{0} \tag{22}$$

where the upper index $[\cdot]$ denotes the maximum possible degree of the respective polynomial $s_{ij}(\hat{r}_x)$.

Now, as in the hidden variable resultant method mentioned previously, we can find an up to degree 8 polynomial in $\hat{r}_x$:

$$\det(\mathbf{M}(\hat{r}_x)) = 0 \tag{23}$$

The unknown $\hat{r}_x$ has at most 8 solutions and can be computed as the eigenvalues of the companion matrix of $\det(\mathbf{M}(\hat{r}_x))$. Then, the corresponding solutions for the unknowns $\hat{r}_y$, $\hat{r}_z$ can be obtained by performing SVD after substituting the particular solutions for $\hat{r}_x$ into $\mathbf{M}(\hat{r}_x)$.

4.3. 2pt calibration method with an unknown focal length for the pure rotation case

In this case, we assume that we have a camera equipped with an IMU with known intrinsic camera parameters except for an unknown common focal length. This is a typical case encountered in practice. For example, it is often practical to assume that the principal point and aspect ratio can be considered as fixed and known for a certain camera [33], the focal length of camera is constant across multiple views.

Brown *et al.*[29] have presented a solution to the problem of estimating rotation and the focal length from two images in the same scene undergoing pure rotation by using two point correspondences. Inspired by Brown *et al.*[29], we firstly compute the focal length f using two point correspondences and normalize image coordinates using the

focal length. Then, we use the Gröbner basis method in Section 4.2.1 or 3Q3 method in Section 4.2.2 to calibrate the rotational alignment between the IMU and the camera.

4.4. Non-linear parameter optimization

Using the 3pt calibration method for the general motion case or the 1.5pt calibration method for the pure rotation case, $\mathbf{R}_{calib}$, $\mathbf{t}_{ij}$ and the corresponding inliers can be obtained for each image pair, leading to N_p inliers in M image pairs, whereby each image pair is referenced by p. Note that all translation parameters $\mathbf{t}_{ij}$ are $\mathbf{0}$ in the pure rotation case. In the optimization step, the translation parameters of the M image pairs are fixed, and the rotation parameters $\mathbf{R}_{calib}$ between the IMU and the camera are optimized using all the inliers. The cost function which minimizes the total transfer errors is as follows:

$$\begin{aligned} \varepsilon &= \min_{\bar{\mathbf{R}}} \sum_{p=1}^{M} \sum_{k=1}^{N_p} \left\| \mathbf{x}_j^k - \mathbf{H}_p \mathbf{x}_i^k \right\| \\ &= \min_{\bar{\mathbf{R}}} \sum_{p=1}^{M} \sum_{k=1}^{N_p} \left\| \mathbf{x}_j^k - g(\bar{\mathbf{R}}, \mathbf{t}_{ij}^p, \mathbf{R}_{imu}^p) \mathbf{x}_i^k \right\|, \end{aligned} \tag{24}$$

where $\bar{\mathbf{R}}$ is the three-parameter rotation estimate used for optimization. For initialization, we set it to the mean or median angles computed from the M calibration results obtained in the previous step. k is the index of the inliers within each image pair p, which is composed of views i and j. $\mathbf{x}_i^k = [x_i^k, y_i^k, 1]^T$ and $\mathbf{x}_j^k = [x_j^k, y_j^k, 1]^T$ are the homogeneous image coordinates of the inlier k, with a unit of pixel. $\mathbf{t}_{ij}^p$ is the translation vector in image pair p, and $\mathbf{R}_{imu}^p$ denotes the IMU rotation matrices of views i and j. The homography $g(\bar{\mathbf{R}}, \mathbf{t}_{ij}^p, \mathbf{R}_{imu}^p)$ is the transformation model, which transfers the homogeneous image coordinate $\mathbf{x}_i$ in view i to the corresponding image coordinate $\mathbf{x}_j$ in view j.

Using 2pt calibration method with the unknown focal length for the pure rotation case, the focal length f and the rotation parameters $\mathbf{R}_{calib}$ between the IMU and the camera are optimized together using all the inliers. The cost function which minimizes the total transfer errors is as follows:

$$\varepsilon = \min_{(\bar{f}, \bar{\mathbf{R}})} \sum_{p=1}^{M} \sum_{k=1}^{N_p} \left\| \mathbf{x}_j^k - g(\bar{f}, \bar{\mathbf{R}}, \mathbf{R}_{imu}^p) \mathbf{x}_i^k \right\|, \tag{25}$$

where $\bar{f}$ and $\bar{\mathbf{R}}$ are the parameters to optimize. For initialization, we also set $\bar{f}$ and $\bar{\mathbf{R}}$ to the mean or median values computed from the M calibration results obtained in the previous step. The definitions of parameters $\mathbf{x}_i^k$, $\mathbf{x}_j^k$ and $\mathbf{R}_{imu}^p$ are the same as in Eq. 24, please refer to Eq. 24.

The Cauchy function is used to create a robust cost function in the optimization process, to reduce the influence of outliers that may still be present.

$$\rho(\varepsilon) = \frac{\sigma^2}{2} \log(1 + \frac{\varepsilon^2}{\sigma^2}). \tag{26}$$

We set the σ parameter of the Cauchy function to 2 pixels, which is similar to the inlier threshold of the RANSAC loop, which is also 2 pixels.

5. Experiments

We validated the performance of the proposed IMU-camera calibration methods using both synthetic and real scene data, including the 3pt calibration method for the general motion case (3pt), the 1.5pt calibration method for the pure rotation case (1.5pt-GB and 1.5pt-3Q3) and the 2pt calibration method with the unknown focal length for the pure rotation case (2pt-GB and 2pt-3Q3).

In all of the experiments, we compared the rotational component between the IMU and the camera (in Euler angles) and compared the relative translation between views i and j separately. The used error measure compares the angle difference between the true rotation and estimated rotation. Since the estimated translation between views i and j is only known up to scale, we compare the angle difference between the true translation and estimated translation. The errors are computed as follows:

- Rotation error: $\xi_{\mathbf{R}} = \arccos((Tr(\mathbf{R}_{gt}\mathbf{R}_{calib}^T) - 1)/2)$
- Translation error: $\xi_{\mathbf{t}} = \arccos((\mathbf{t}_{gt}^T\mathbf{t})/(\|\mathbf{t}_{gt}\| \, \|\mathbf{t}\|))$

$\mathbf{R}_{gt}$, $\mathbf{t}_{gt}$ denote the ground-truth transformation and $\mathbf{R}_{calib}$, $\mathbf{t}$ are the corresponding estimated transformations.

5.1. Experiments with synthetic data

5.1.1. Accuracy with increasing rotation

We evaluate our approach with respect to increasing amounts of remaining rotation, as we approximate the remaining rotation matrix to the first-order and truncate the higher-order terms. For this experiment, normalized image points are generated randomly and point matches are computed by the ground truth homography. The number of independent trials is 10000, and three approximate installation angles between the IMU and the camera are chosen randomly, from -180° to 180°. We set three approximate angles between the IMU and the camera as known and use this approximate rotation matrix to rotate the image features first. The three remaining angles between the IMU and the camera are then increased from 0° to 10° in steps of 1°. We assess the rotation and translation error in three different ways: $Mean$ denotes the mean value of the errors, $Median$ denotes the median value of the errors and $RMSE$ denotes the root mean square error of the errors.

We report the results on the data points within the first interval of a 5-quantile partitioning[1] (Quintile) of 10000 trials. The errors for the rotational component and translation are reported in Figures 3. There is no significant difference among $Mean$, $Median$ and $RMSE$, and when the three approximate installation angles between the IMU and the camera are known, the errors increase slowly as the remaining rotation magnitude increases. It shows that our methods are numerically stable and do not show a significant loss in accuracy even at the maximum magnitude for the remaining rotation angles up to 10°. From the Figures 3, we can also see that the 3pt calibration method for the general motion case returns slightly more accurate estimates than the 1.5pt calibration method for the pure rotation case. One reason for this is that Eq. 15 is only composed of a rotation matrix, so the pure rotation case is generally more sensitive to the rotation magnitude. Notice that the 1.5pt-GB and 1.5pt-3Q3 methods have similar accuracy with increasing magnitudes of rotation. The 2pt calibration method with the unknown focal length has not been performed, because the computation of

[1] k-quantiles divide an ordered data set into k regular intervals

focal length is not influenced by the rotation magnitude between the IMU and the camera, so the rotational component error for the 2pt method is as same as for the 1.5pt method.

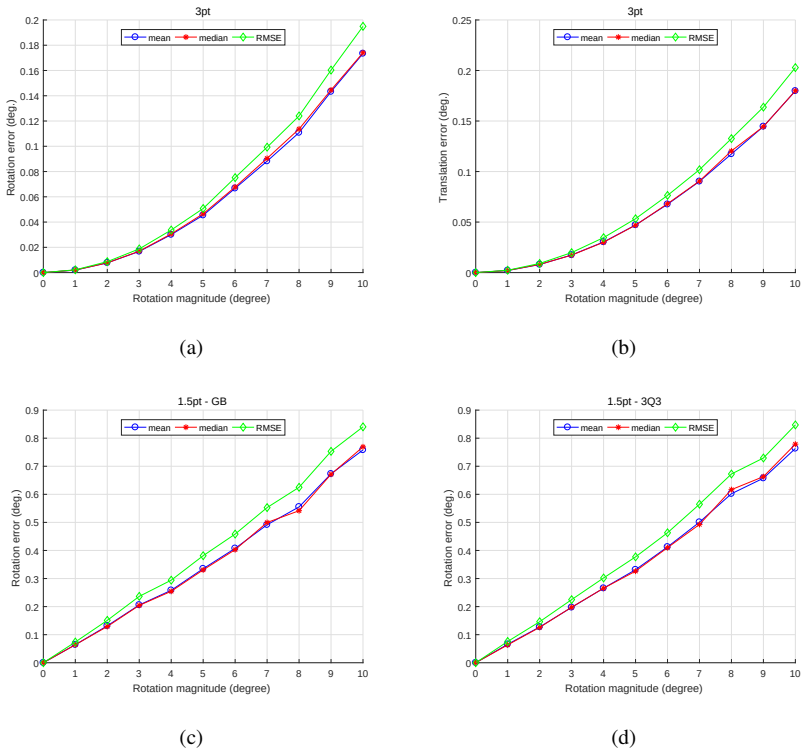


Figure 3: Mean, Median and RMSE for the rotational component and translation with increasing magnitudes of remaining rotation. No noise is added to the observations. (a) and (b) are the rotational component and translation errors for the 3pt method, respectively. (c) and (d) are the rotational component errors for the 1.5pt-GB and 1.5pt-3Q3 methods, respectively.

5.1.2. Accuracy with increasing image noise

We synthesize a pinhole camera with zero skew and an unit aspect ratio that has a resolution of 800×640 pixels. The principle point is assumed to be at the image center. A different level of Gaussian noise with a standard deviation ranging from 0 to 2 pixels is then added to the image feature observations. The approximate installation angles between the IMU and the camera are set to $(180°, 0°, -90°)$, while keeping the

remaining rotation angles constant at $(1^\circ, 1^\circ, -1^\circ)$. The focal length is chosen as 600 pixels, so that one pixel corresponds to about 0.1°. At each noise level, 10000 independent trials are conducted, and for each test, we select 3 image features randomly for the 3pt method, or 2 image features randomly for the 1.5pt and 2pt methods. The errors for the rotational component and translation are reported in Figure 4. As in the previous experiment, we report the results on the data points within the first one interval of a 5-quantile partitioning. The accuracy of our method is observed to decrease almost linearly with the increase in image noise. We can clearly see that the 3pt calibration method for the general motion case produces much better results than the 1.5pt and 2pt calibration methods for the pure rotation case. For the pure rotation case, no matter what the focal length error or rotational component error, we do not find any difference in accuracy between the GB method and 3Q3 method. Figure 4(e) and (g) are significantly different in terms of the $Mean$, $Median$ and $RMSE$ of the focal length error, and the $RMSE$ is quite shaky, because we generate 2 image points randomly to compute the focal length for each test, and the accuracy of focal length is influenced by the distribution of the image points.

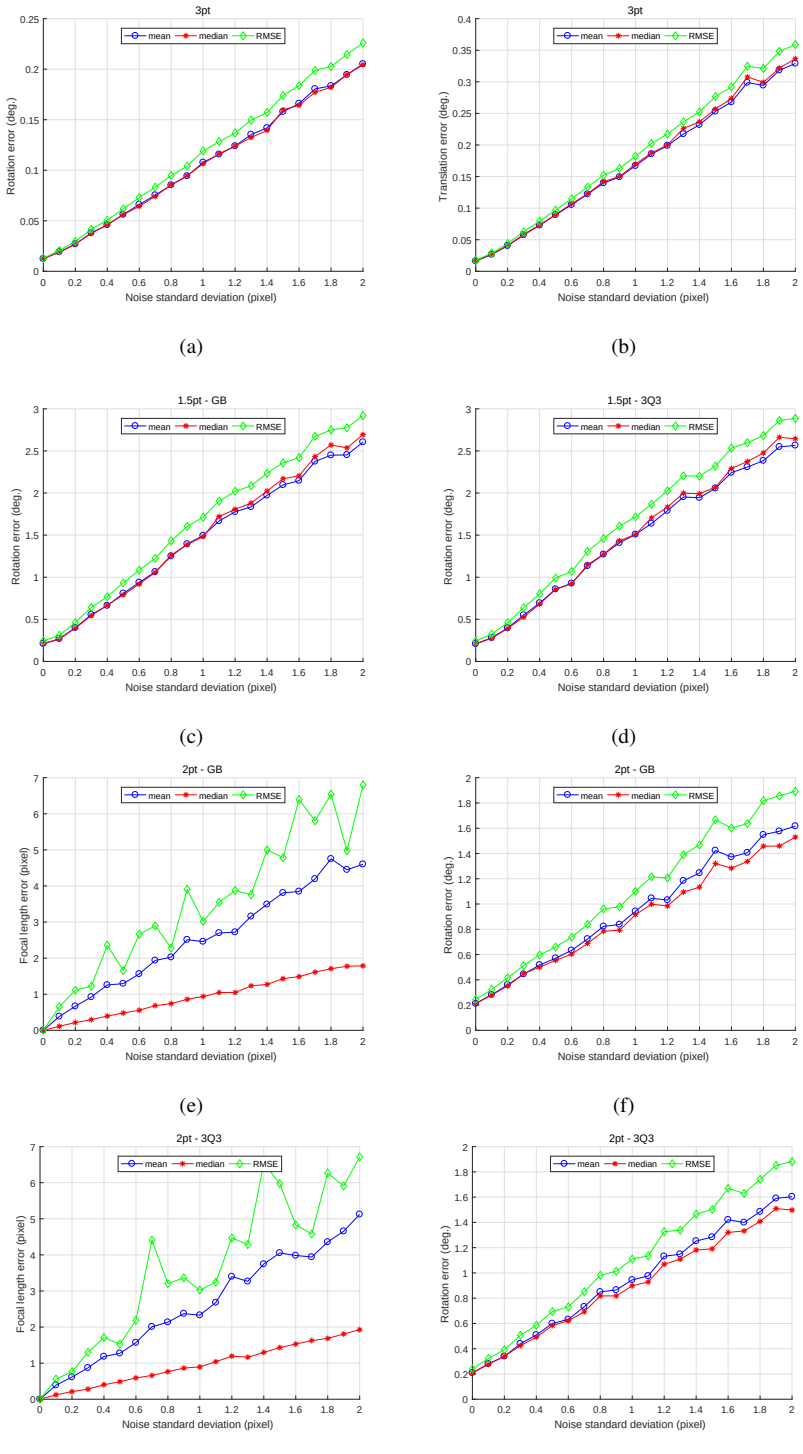


Figure 4: Mean, Median and RMSE for the rotational component and translation with increasing image noise, with the approximate installation angles $(180°, 0°, -90°)$ and the remaining rotation angles $(1°, 1°, -1°)$ between the IMU and the camera. (a) and (b) are the rotational component and translation errors for the 3pt method, respectively. (c) and (d) are the rotational component errors for the 1.5pt-GB and 1.5pt-3Q3 methods, respectively. (e, f), and (g, h) are the estimated focal length and rotational component errors for the 2pt-GB and 2pt-3Q3 methods, respectively.

5.2. Real scene data experiment

Our real image data sets consist of a data set from a mobile robot, a data set acquired with a common smartphone (SONY LT22i) and a data set from a micro aerial vehicle (MAV). For each of these data sets, we show the results of a detailed analysis and compare these with those obtained using state-of-the-art methods. The robotic scenario and smartphone data set are used to evaluate the 3pt method. Using the smartphone data set, we also demonstrate that the proposed method can be used with a small number of images under challenging conditions. All the proposed calibration methods were also evaluated with the MAV data set. We compared our methods to state-of-the-art methods that can handle the same input data, which are small wide baseline image data sets without calibration targets. Methods which need specific calibration targets in the images and require video data were not used in our comparison *e.g.* Crisp [17] and Kalibr [19, 37].

For each data set, we consider feasible image pairs for image matching. For each image pair, features matches are created using SURF feature matching [38], and the calibration parameters are estimated using our method within a RANSAC loop [10]. We use an inlier threshold of 2 pixels and a fixed number of 100 iterations for RANSAC. All inliers of all the image pairs are stored for the subsequent optimization step. Considering that different rotational estimates have been computed for each image pair, we choose the median and mean angle values of the rotations of all image pairs as the initial values for non-linear parameter optimization, respectively. However, the optimization converged to the same result for both initializations in all experiments. The intrinsic parameters of the cameras were obtained in advance, except for the 2pt method. Finally, we also obtained a comparison to the ground truth by using a calibration target for all our methods.

5.2.1. Real data from the Vicon data set

The Vicon data set has been acquired with a perspective camera mounted on a mobile robot. The camera is synchronized with a Vicon motion capture system consisting of 22 tracking cameras. Vicon markers are attached to the camera mount and the pose is tracked by the Vicon system. In this experiment, the Vicon poses are used as IMU data.

Furthermore, for the comparisons, we scale our translation directions with the metric scale obtained from the Vicon system. The approximate installation angles between the IMU and the camera are $(180°, 0°, 0°)$. The camera is typically looking towards the ground, and 219 images of 1624×1234 pixels are captured. To obtain expressive results, we compare the 3pt method to a range of reference implementations: Tsai89 [9], Park94 [8], Horaud95[2] [6], Daniilidis99 [5] and Heller16 [24]. As the methods require the prior knowledge of the camera poses or the image correspondences, the open source SfM pipeline COLMAP [39] is used to recover the poses of images. The metric scale is recovered by using the data from the Vicon system. The image poses are taken as input parameters for the hand-eye calibration methods Tsai89 , Park94, Horaud95 and Daniilidis99, while the inlier matches determined by COLMAP are used in Heller16.

Table 1 shows the calibration results obtained by the computations using all the methods. There is no ground truth for the rotational component between the IMU and the camera, so we cannot assess the accuracy quantitatively. As can be seen, Park94, Horaud95 and our approach are close to the installation angles. We were not able to produce a result with Heller16 for a data set of this size, Tsai89 and Daniilidis99 have a significant deviation from the actual installation angles. The hand-eye calibration methods typically rely on accurate and outlier free pose estimates, and small inaccuracies, typical to SfM pipelines, will already produce large deviations in the calibration results. Although the methods of Tsai89, Park94, Horaud95 and Daniilidis99 solve for the same equation system, they use different parameterization for the transformation parameters, leading to different results. Heller16 requires the use of outlier free feature tracks for each image pair to construct the optimization task. As these methods have not adopted any procedures to cope with outliers, either in the transformations or in the feature matches, the accuracy is inevitably influenced by such outliers. In contrast, our 3pt method uses three point correspondences directly and performs RANSAC as a framework for robust estimation. This experiment successfully demonstrates the practicability of our proposed 3pt method. The histogram of inlier transfer errors for the

[2]We use the first method to solve for the rotational and translational components separately.

Method	Calibration results (degree)
Approximate installation angle	(180.0, 0.0, 0.0)
Tsai89	(-8.91, -56.05, 12.21)
Park94	(180.73, -0.84, -2.33)
Horaud95	(180.72, -0.79, -2.30)
Daniilidis99	(33.46, -34.00, -176.81)
Heller16	\
Our method	Mean: (181.35, 0.16, -0.17) Median: (181.18, 0.61, -0.16) Optimization: (181.40, 1.74, 1.10)

Table 1: The calibration results for the Vicon data set. For our method, non-linear parameter optimization yields the same final calibration result when initialized with either the mean or the median values, so only one optimization calibration result is shown here. Tsai89 and Daniilidis99 show strong deviations in this experiment.

Vicon data set is shown in Figure 5. The inlier transfer error is computed from the individual terms in Equation 24.

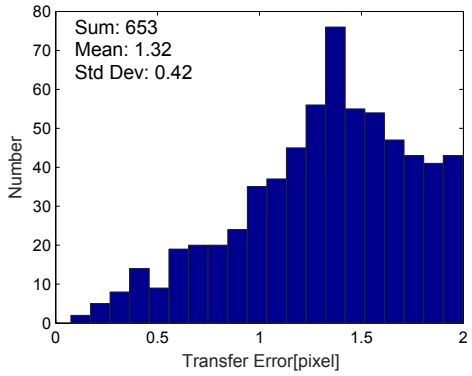


Figure 5: Histogram of inlier transfer errors for the Vicon data set using the 3pt method. In all of experiments, the labels “Sum“, “Mean“ and “Std Dev“ stand for total number of inlier, mean and standard deviation of inlier transfer errors, respectively.

Method	Calibration results (degree)
Approximate installation angle	(180.0, 0.0, 180.0)
Tsai89	(-32.06, 25.07, -67.71)
Park94	(184.00, -0.39, 187.54)
Horaud95	(194.49, -7.08, -31.53)
Our method	Mean: (181.16, 0.82, 179.43)
	Median: (180.25, 0.39, 179.69)
	Optimization: (179.47, 3.26, 180.23)

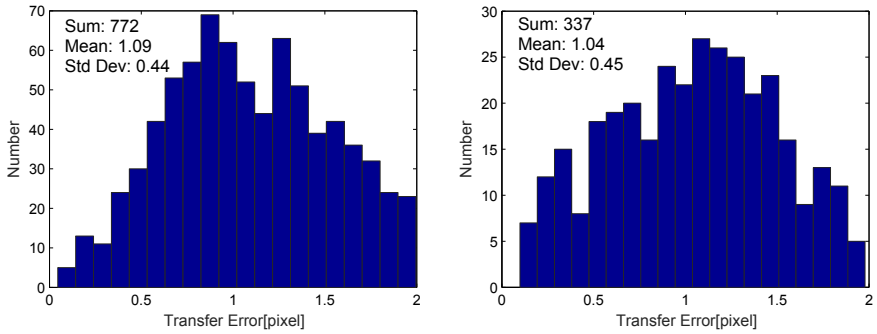
Table 2: The calibration results for the Sony data set using all the 42 images. Tsai89 and Horaud95 cannot produce correct results.

5.2.2. Real data from the SONY LT22i smartphone

To demonstrate that the 3pt method also works on currently-available consumer smartphones, we tested it with the SONY LT22i equipped with a camera and an IMU.

We determined the approximate installation angles of the SONY LT22i to be $(180°, 0°, 180°)$. 42 images of 3264×2448 pixels are captured by its rear camera. Due to the lack of translation information of the smartphone, Daniilidis99 and Heller16 cannot be tested for comparison.

Like the Vicon data set, the image poses are computed using COLMAP. Table 2 shows the calibration results computed by the 3pt method and the other hand-eye calibration methods. As can be seen, the results yielded by Tsai89 and Horaud95 significantly deviate from the actual installation angles. Park94 and our 3pt method, in comparison, yield results that are close to the installation angles. This shows that our 3pt method is effective for this scenario as well. The histogram of inlier transfer errors for the Sony data set is shown in Figure 6(a). The orthophotos of the images rectified using the calibration results are shown in Figure 7.



(a) 42 images (b) 8 images

Figure 6: Histogram of inlier transfer errors for the Sony data set.

Figure 7: Orthophotos created using the calibration results. Three representative images are shown as original images (left) and orthophotos (right), and the size of the orthophotos is determined by the maximum value of image boundaries. Obviously, the edges of the magazines are perpendicular in the orthophotos.

Method	Calibration results (degree)
Approximate installation angle	(180.0, 0.0, 180.0)
Tsai89	(-0.57, -3.35, 206.68)
Park94	(-53.39, -12.82, 189.70)
Horaud95	(1.17, -12.47, 190.74)
Our method	Mean: (178.76, 1.31, 176.71)
	Median: (178.51, 0.04, 179.11)
	Optimization: (181.28, -0.78, 178.30)

Table 3: The calibration results for the Sony data set using only a subset of 8 images. Only our 3pt method produces a correct result.

To test the robustness of our 3pt method, we perform an experiment under the challenging condition of using only a small number of images. We only take 8 images of the data set for a calibration experiment. The calibration results in Table 3 show that only our 3pt method works effectively for this challenging data set. The histogram of inlier transfer errors is shown in Figure 6(b).

Our 3pt method computes the camera motion and the IMU-camera calibration simultaneously. This allows us to visualize the camera motion. We align the pose of one camera of the data set with an estimate from COLMAP and transfer the scale from COLMAP to our results. The camera motion of the challenging data set recovered by our 3pt method is shown in Figure 8. Compared with COLMAP, the rotation and translation differences are shown in Table 4. Our method achieves comparable reconstruction results as SfM pipelines, while the rotational component between the IMU and the camera is computed as well. It should be noted that most methods require performing SfM on the images to create the input data, while ours can also be used to compute the camera motion.

Images	1	2	3	4	5	6	7
$\xi_{\mathbf{R}}$(deg.)	2.52	0.93	0.56	0.64	1.65	1.40	1.41
$\xi_{\mathbf{t}}$(deg.)	1.76	0.77	0.67	0.73	2.20	1.40	1.13

Table 4: The rotation and translation differences of the camera poses between our 3pt method and COLMAP for the Sony 8 images data set.

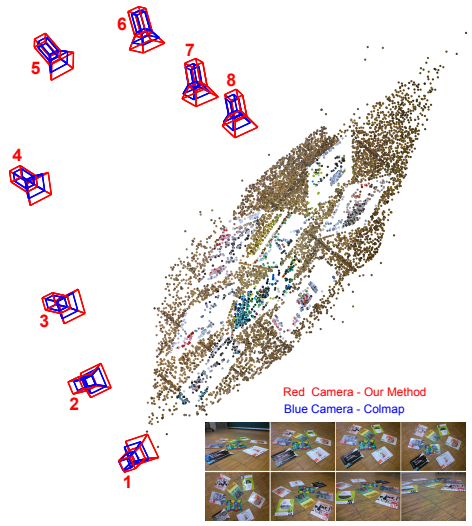


Figure 8: Camera poses for the 8 images Sony data set. The camera motion recovered by our 3pt method and COLMAP. The 8 images used are shown in the lower right corner.

5.2.3. Real data from the MAV

To demonstrate the 3pt (both rotation and translation for the camera), the 1.5pt (pure rotation for the camera) and the 2pt (pure rotation for the camera with unknown focal length) calibration methods in a realistic scenario we have collected two data sets with our Pixhawk drone, see Figure 9. The data sets under general motion and pure rotation have been obtained by conducting the experiments in a room equipped with a motion capture system consisting of 10 cameras. Markers are attached to the camera mount and the pose is tracked by the motion capture system. In this experiment, the marker poses are used as IMU data in the experiments. The approxi-

mate installation angles between the IMU and the RGB camera are $(113°, 0°, 90°)$, which come from the design of the 3D printed mount. The offset of the RGB camera and the depth camera has been calibrated beforehand, which is a pure translation $(0.000, -0.020, 0.000)m$. The camera is typically looking towards the ground, and the resolution of images is 640×480 pixels. The intrinsic matrix of RGB camera has been calibrated using the popular Bouguet toolbox [40]: $\mathbf{K} = [536.29461 \quad 0 \quad 317.76263; \; 0 \quad 536.18547 \quad 238.81011; \; 0 \quad 0 \quad 1]$, and the lens distortion is $(0.04234, -0.12481, -0.00040, -0.00029, 0.00000)$. 23 images under general motion are captured for the 3pt calibration methods, and 81 images under pure rotation are captured for the 1.5pt and the 2pt calibration methods.

(a) Pixhawk drone (b) Sample image

Figure 9: MAV data set. (a). Pixhawk drone capturing image. (b). Sample image captured by the Pixhawk drone.

In Table 5, we show the calibration results of the different calibration algorithms. All calibration results are similar. All our methods (the 1.5pt, 2pt, GB and 3Q3 methods) have quite consistent calibration results. The focal length estimation using the 2pt method is also quite accurate compared to the calibration results obtained with the Bouguet toolbox.

5.2.4. Accuracy evaluation

In this final experiment, we continue to use the MAV to acquire a data set of images for a calibration target to evaluate the accuracy of the calibration results as compared to the ground truth (see Table 5). The calibration target consists of a checkerboard which

Method	Calibration results (degree)/ Focal length (pixel)
Approximate installation angle	(113.0, 0.0, 90.0)
3pt	(114.1171, 1.3155, 88.6784)
1.5pt-GB	(114.4211, 1.2609, 88.7395)
1.5pt-3Q3	(114.4241, 1.2845, 88.74310)
2pt-GB	(114.5166, 2.6220, 88.4692), f = 539.4631
2pt-3Q3	(114.3481, 2.4268, 88.6699), f = 543.5531

Table 5: The calibration results for the MAV data set. Only the calibration results after non-linear parameter optimization are shown in this table.

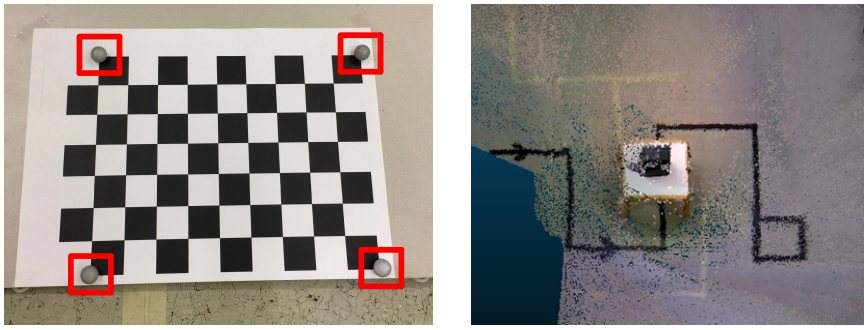
is horizontally placed on the ground floor. Four motion tracking markers are placed onto the corners of the checkerboard, see Figure 10(a). The size of each checker is $3.5cm \times 3.5cm$. The coordinates of the four markers are measured using the motion capture system, which are the coordinates of the four outmost corners of the checkerboard. The precise coordinates on the calibration target are estimated, taking into account the radius of the markers, which is $0.85cm$. Then, all the remaining coordinates of the checkerboard corners are computed from the measured outmost corner coordinates.

We randomly take 49 images around the checkerboard at the distance of $1m$, and we use OPnP algorithm [41] to compute the pose of each image. Combined with the corresponding IMU data for each image, we can compute the relationship between the IMU and the camera directly and accept the mean of 49 images as the ground truth. The relationship between the IMU and the camera is as follows: the rotational component is $(114.1497^\circ,\ 1.1152^\circ, 88.7120^\circ)$ and translational component is $(0.0316, 0.0222, -0.0638)m$.

For comparison, we fix the translational component between the IMU and the camera as $(0.0316, 0.0222, -0.0638)m$. Only the rotational component between the IMU and the camera is calibrated using our methods. The reprojection error is used to evaluate the accuracy of our calibration results. The reprojection error is the mean distance between the measured image corners and the reprojection of the 3D corner of the cal-

ibration target using our calibration results. The results of the different methods are shown in Table 6. When we use the approximate installation angles between the IMU and the RGB camera $(113°, 0°, 90°)$ which come from the design of the 3D printed mount, the reprojection error is 11.5408 pixel. However, the calibration results of our methods produce lower reprojection errors than using the approximate installation angles directly. It means that even though the final estimation is mostly very close to the initial estimation (within 2 degrees), it is still necessary to calibrate the rotational component between the IMU and the camera. The table shows that the 3pt method outperforms other methods in terms of accuracy, it is consistent with the synthetic results that the pure rotation case is more sensitive to rotation magnitude and image noise than general motion case, as showed in Figures 3 and Figure 4. Our GB method and 3Q3 method exhibit similar accuracy in terms of reprojection error. The 1.5pt method performs better than the 2pt method, because the 1.5pt method has been performed with known intrinsic parameters, but the 2pt method has been performed with unknown focal length and lens distortion parameters and cannot handle any image distortion. The 2pt method inevitably results in terms of a loss in accuracy for lenses with distortion.

After performing IMU-camera calibration, we obtain the pose of the RGBD camera directly. To verify the calibration results intuitively, we reconstruct a common scene using the RGBD camera. We take the reconstruction results based on the calibration results of the 3pt method, as example, see Figure 10(b). This experiment successfully demonstrates the practicability of our proposed calibration method.



(a) Checkerboard (b) Reconstruction result

Figure 10: Accuracy evaluation. (a). Checkerboard and four markers in red box. (b). Reconstruction result using RGBD camera based on the calibration results of 3pt method.

Calibration results	Ground Truth	3D printer	3pt	1.5pt-GB	1.5pt-3Q3	2pt-GB	2pt-3Q3
Reprojection error (pixel)	1.3495	11.5408	1.5441	2.7785	2.8244	3.6832	3.6494

Table 6: The results of the accuracy evaluation using the checkerboard data set captured by the MAV. The reprojection errors are used to evaluate the accuracy of different calibration methods as shown in Table 5.

6. Conclusion

In this paper, we focused on the rotational alignment of IMU-camera systems. We presented novel minimal case solutions to the IMU-camera calibration problem utilizing a first-order rotation approximation. We formulated this problem as a problem of solving a polynomial equation system derived from homography constraints. This made it possible to derive algorithms that need fewer point correspondences for IMU-camera calibration as compared to state-of-art methods. We derived the solution with minimal point correspondences varying from 1.5 to 3 using Gröbner basis method and analytical solver. By evaluating our algorithm on synthetic and real-world image data sets, we demonstrated that our method is more efficient and numerically stable for IMU-camera calibration compared to state-of-the-art methods.

Future work includes developing applications for smart devices, such as real-time rectification for tilted pictures and image stabilization.

Acknowledgements

We thank Pascal Vasseur and Rmi Boutteau for making their Vicon data set available. We also thank Werner Alexander Isop for helping us collect the MAV data set. The research was supported by the National Basic Research Program of China (973 Program) (No.2013CB733101), Scientific Research Program of National University of Defense Technology (No.ZK16-03-37) and National Natural Science Foundation of China (No.11332012).

References

[1] L. Kneip, M. Chli, R. Siegwart, Robust real-time visual odometry with a single camera and an imu, in: BMVC, BMVC, 2011, pp. 1–11.

[2] C. Ham, S. Lucey, S. Singh, Hand Waving Away Scale, Springer International Publishing, Cham, 2014, pp. 279–293. doi:10.1007/978-3-319-10593-2_19.
URL http://dx.doi.org/10.1007/978-3-319-10593-2_19

[3] O. Saurer, P. Vasseur, R. Boutteau, C. Demonceaux, M. Pollefeys, F. Fraundorfer, Homography based egomotion estimation with a common direction, IEEE Transactions on Pattern Analysis and Machine Intelligence PP (99) (2016) 1–1. doi:10.1109/TPAMI.2016.2545663.

[4] F. Fraundorfer, P. Tanskanen, M. Pollefeys, A Minimal Case Solution to the Calibrated Relative Pose Problem for the Case of Two Known Orientation Angles, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 269–282. doi:10.1007/978-3-642-15561-1_20.
URL http://dx.doi.org/10.1007/978-3-642-15561-1_20

[5] K. Daniilidis, Hand-eye calibration using dual quaternions, The International Journal of Robotics Research 18 (3) (1999) 286–298. arXiv:http://ijr.sagepub.com/content/18/3/286.full.pdf+html, doi:10.1177/02783649922066213.
URL http://ijr.sagepub.com/content/18/3/286.abstract

[6] R. Horaud, F. Dornaika, Hand-eye Calibration, International Journal of Robotics Research 14 (3) (1995) 195–210. doi:10.1177/027836499501400301.
URL https://hal.inria.fr/inria-00590039

[7] Z. Kukelova, J. Heller, T. Pajdla, Hand-Eye Calibration without Hand Orientation Measurement Using Minimal Solution, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 576–589. doi:10.1007/978-3-642-37447-0_44.
URL http://dx.doi.org/10.1007/978-3-642-37447-0_44

[8] F. C. Park, B. J. Martin, Robot sensor calibration: solving ax=xb on the euclidean group, IEEE Transactions on Robotics and Automation 10 (5) (1994) 717–721. doi:10.1109/70.326576.

[9] R. Y. Tsai, R. K. Lenz, A new technique for fully autonomous and efficient 3d robotics hand/eye calibration, IEEE Transactions on Robotics and Automation 5 (3) (1989) 345–358. doi:10.1109/70.34770.

[10] M. A. Fischler, R. C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395. doi:10.1145/358669.358692.
URL http://doi.acm.org/10.1145/358669.358692

[11] J. Kelly, G. S. Sukhatme, Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration, The International Journal of Robotics Research 30 (1) (2011) 56–79.

[12] F. M. Mirzaei, S. I. Roumeliotis, A kalman filter-based algorithm for imu-camera calibration: Observability analysis and performance evaluation, IEEE Transac-

tions on Robotics 24 (5) (2008) 1143–1156. doi:10.1109/TRO.2008.2004486.

[13] S. Weiss, M. W. Achtelik, M. Chli, R. Siegwart, Versatile distributed pose estimation and sensor self-calibration for an autonomous mav, in: Robotics and Automation (ICRA), 2012 IEEE International Conference on, 2012, pp. 31–38. doi:10.1109/ICRA.2012.6225002.

[14] S. Lovegrove, A. Patron-Perez, G. Sibley, Spline fusion: A continuous-time representation for visual-inertial fusion with application to rolling shutter cameras, in: British Machine Vision Conference, 2013, pp. 93.1–93.11.

[15] C. Jia, B. L. Evans, Online calibration and synchronization of cellphone camera and gyroscope, in: Global Conference on Signal and Information Processing, 2013, pp. 731–734.

[16] M. Li, H. Yu, X. Zheng, A. I. Mourikis, High-fidelity sensor modeling and self-calibration in vision-aided inertial navigation, in: IEEE International Conference on Robotics and Automation, 2014, pp. 409–416.

[17] H. Ovren, P. E. Forssn, Gyroscope-based video stabilisation with auto-calibration 2015 (2015) 2090–2097.

[18] J. Lobo, J. Dias, Relative pose calibration between visual and inertial sensors, The International Journal of Robotics Research 26 (6) (2007) 561–575.

[19] P. Furgale, J. Rehder, R. Siegwart, Unified temporal and spatial calibration for multi-sensor systems, in: Ieee/rsj International Conference on Intelligent Robots and Systems, 2013, pp. 1280–1286.

[20] Z. Q. Zhang, Cameras and inertial/magnetic sensor units alignment calibration, IEEE Transactions on Instrumentation and Measurement 65 (6) (2016) 1495–1502. doi:10.1109/TIM.2016.2518418.

[21] H. Zhuang, Y. C. Shiu, A noise tolerant algorithm for wrist-mounted robotic sensor calibration with or without sensor orientation measurement, in: Intelligent

Robots and Systems, 1992., Proceedings of the 1992 lEEE/RSJ International Conference on, Vol. 2, 1992, pp. 1095–1100. doi:10.1109/IROS.1992.594526.

[22] T. Ruland, T. Pajdla, L. Krger, Globally optimal hand-eye calibration, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 1035–1042. doi:10.1109/CVPR.2012.6247781.

[23] J. Heller, M. Havlena, T. Pajdla, A branch-and-bound algorithm for globally optimal hand-eye calibration, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 1608–1615. doi:10.1109/CVPR.2012.6247853.

[24] J. Heller, M. Havlena, T. Pajdla, Globally optimal hand-eye calibration using branch-and-bound, IEEE Transactions on Pattern Analysis and Machine Intelligence 38 (5) (2016) 1027–1033. doi:10.1109/TPAMI.2015.2469299.

[25] D. Bender, M. Schikora, J. Sturm, D. Greniers, Ins-camera calibration without ground control points, in: Sensor Data Fusion: Trends, Solutions, Applications (SDF), 2014, 2014, pp. 1–6. doi:10.1109/SDF.2014.6954719.

[26] Y. Seo, Y.-J. Choi, S. W. Lee, A branch-and-bound algorithm for globally optimal calibration of a camera-and-rotation-sensor system, in: 2009 IEEE 12th International Conference on Computer Vision, 2009, pp. 1173–1178. doi:10.1109/ICCV.2009.5459343.

[27] M. Hwangbo, J.-S. Kim, T. Kanade, Gyro-aided feature tracking for a moving camera: fusion, auto-calibration and gpu implementation, The International Journal of Robotics Research 30 (14) (2011) 1755–1774.

[28] A. Karpenko, D. Jacobs, J. Baek, M. Levoy, Digital video stabilization and rolling shutter correction using gyroscopes, CSTR 1 (2011) 2.

[29] M. Brown, R. I. Hartley, D. Nister, Minimal solutions for panoramic stitching, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.

[30] G. H. Golub, C. F. Van Loan, Matrix computations, Mathematical Gazette 47 (5 Series II) (1996) 392–396.

[31] L. Dorst, First order error propagation of the procrustes method for 3d attitude estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2) (2005) 221–9.

[32] J. Ventura, C. Arth, V. Lepetit, An efficient minimal solution for multi-camera motion, in: IEEE International Conference on Computer Vision, 2015, pp. 747–755.

[33] R. Hartley, A. Zisserman, Multiple view geometry in computer vision (2003).

[34] C. David A, L. John, O. Donal, Ideals varieties and algorithms: an introduction to computational algebraic geometry and commutative algebra (2007).

[35] Z. Kukelova, M. Bujnak, T. Pajdla, Automatic Generator of Minimal Problem Solvers, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 302–315. doi:10.1007/978-3-540-88690-7_23.
URL http://dx.doi.org/10.1007/978-3-540-88690-7_23

[36] Z. Kukelova, J. Heller, A. Fitzgibbon, Efficient intersection of three quadrics and applications in computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1799–1808.

[37] P. Furgale, T. D. Barfoot, G. Sibley, Continuous-time batch estimation using temporal basis functions, in: IEEE International Conference on Robotics and Automation, 2012, pp. 2088–2095.

[38] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, Speeded-up robust features (surf), Computer Vision and Image Understanding 110 (3) (2008) 346 – 359, similarity Matching in Computer Vision and Multimedia. doi:http://dx.doi.org/10.1016/j.cviu.2007.09.014.
URL http://www.sciencedirect.com/science/article/pii/S1077314207001555

[39] J. L. Schönberger, J.-M. Frahm, Structure-from-motion revisited, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[40] J.-Y. Bouguet, Camera calibration toolbox for matlab.$http : //www.vision.caltech.edu/bouguetj/calib\_doc.$

[41] Y. Zheng, Y. Kuang, S. Sugimoto, K. Astrom, M. Okutomi, Revisiting the pnp problem: A fast, general and optimal solution, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 2344–2351.