Trust Your IMU: Consequences of Ignoring the IMU Drift

Marcus Valtonen Örnhag[1], Patrik Persson[1], Mårten Wadenbäck[2], Kalle Åström[1], Anders Heyden[1]

[1]Centre for Mathematical Sciences   [2]Department of Electrical Engineering
Lund University   Linköping University

marten.wadenback@liu.se

Abstract

In this paper, we argue that modern pre-integration methods for inertial measurement units (IMUs) are accurate enough to ignore the drift for short time intervals. This allows us to consider a simplified camera model, which in turn admits further intrinsic calibration. We develop the first-ever solver to jointly solve the relative pose problem with unknown and equal focal length and radial distortion profile while utilizing the IMU data. Furthermore, we show significant speed-up compared to state-of-the-art algorithms, with small or negligible loss in accuracy for partially calibrated setups.

The proposed algorithms are tested on both synthetic and real data, where the latter is focused on navigation using unmanned aerial vehicles (UAVs). We evaluate the proposed solvers on different commercially available low-cost UAVs, and demonstrate that the novel assumption on IMU drift is feasible in real-life applications. The extended intrinsic auto-calibration enables us to use distorted input images, making tedious calibration processes obsolete, compared to current state-of-the-art methods. Code available at: https://github.com/marcusvaltonen/DronePoseLib.[1]

1. Introduction

A popular approach in Simultaneous Localization and Mapping (SLAM) is to fuse various sensor data to increase the performance of the system. A common pair of sensors to combine is a camera and an IMU. Systems of this kind are labeled as visual-inertial odometry, and this specific sensor combination is often found on consumer devices, such as smartphones and UAVs. As is well-known,

Figure 1. The proposed 4-point solver is able to accurately perform radial distortion auto-calibration for focal length and motion parameters. This is a novel case in the literature and is primarily made feasible by a clever assumption—to use the complete rotation estimate from pre-integrated IMU data. We perform experiments with two different UAVs in various difficult scenarios, demonstrating that this assumption comes with a vast array of benefits, *e.g.* simpler equations, faster solvers, and little to no loss in accuracy compared to other state-of-the-art methods.

the projective relationship between two cameras manifests itself in the fundamental matrix, independent of the scene geometry. When auxiliary data are known (*e.g.* IMU data) the number of degrees of freedom decreases and the corresponding fundamental matrix is constrained, which enables one to compute it with fewer point correspondences. This potentially reduces the impact of noise; however, such algebraic constraints can be highly non-trivial to incorporate in a solver. Popular methods to handle the resulting polynomial systems of equations use theory from algebraic geometry, *e.g.* the action matrix method [3], turning the problem

into a generalized eigenvalue problem (GEP) [19, 21], and resultant based methods [1]. Regardless of which method is used, there is still much work in terms of parameterizing the original problem, as different approaches may yield completely different results [24]. For the resulting solver to be of any practical use, it must also be numerically stable, which adds further considerations to the design.

Since modern image sensors often have square-shaped pixels and the lens is sufficiently aligned such that the principal point coincides with the optical center, a feasible assumption is to use partially calibrated cameras, where the only unknown intrinsic parameter is the focal length. This specific assumption has proven useful in several different real scenarios including relative pose estimation [6, 12, 14, 18, 23] and absolute pose estimation [26, 42].

Although it often comes at the cost of introducing distortion, having a wide field of view is desirable in many applications. When working with visual odometry, it is, therefore, a standard procedure to correct for these undesirable distortion artifacts, which often requires a specific calibration setup, typically involving a checkerboard pattern. By incorporating a distortion model, as well as focal length, together with the motion model, one may omit such procedures altogether; however, due to the difficulty of the problem, no fast and robust minimal solver has yet been proposed. The main contributions of this paper are:

- We take advantage of IMU data to estimate the full 3D orientation. Under the assumption that the IMU drift is negligible for short time intervals, the resulting polynomial systems of equations are significantly easier to solve.

- By using this approach, we are able to treat the partially calibrated case with unknown radial distortion profile while incorporating the IMU data, resulting in a fast and reliable solver. This is the only solver to handle this case to date.

- Furthermore, we show a considerable speed-up compared to other state-of-the-art methods, with small or insignificant loss in accuracy, when exploiting the assumption of negligible IMU drift. This benefits low-cost and embedded devices, which constitute the majority of consumer devices where these algorithms are used in practice.

2. Previous Work

2.1. Visual-Inertial Odometry

The calibrated visual-inertial problem of relative pose is well-studied [9, 28, 33, 38] and efficient solvers exist. If we assume that the gravity direction is aligned with the y-axis, the corresponding essential matrix (after alignment) is given by $\boldsymbol{E}_y \sim [\boldsymbol{t}]_\times \boldsymbol{R}_y$, or explicitly,

$$\boldsymbol{E}_y = \begin{bmatrix} -t_y \sin\phi & -t_z & t_y \cos\phi \\ t_z \cos\phi + t_x \sin\phi & 0 & t_z \sin\phi - t_x \cos\phi \\ -t_y \cos\phi & t_x & -t_y \sin\phi \end{bmatrix}, \quad (1)$$
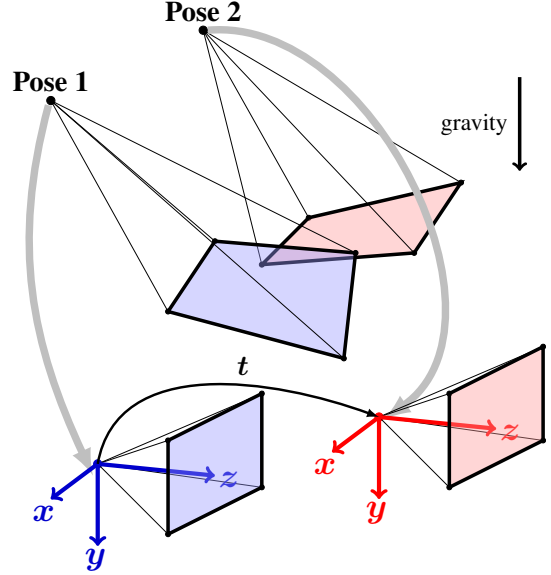

Figure 2. Assume the IMU measurements are accurate, *i.e.* the accelerometer and gyroscope data can be used to accurately estimate the relative orientation between two consecutive views. Then the only unknown extrinsic parameter is the translation vector between the poses.

where

$$\boldsymbol{R}_y = \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{bmatrix}, \quad (2)$$

and $\boldsymbol{t} = (t_x, t_y, t_z)$. This makes it possible to use a parameterization with six elements,

$$\boldsymbol{E}_y = \begin{bmatrix} e_1 & e_2 & e_3 \\ e_4 & 0 & e_5 \\ -e_3 & e_6 & e_1 \end{bmatrix}. \quad (3)$$

Since we have four degrees of freedom (three translation elements and one angle), the elements e_i of (3) are not independent. In fact, one can check that they must obey the (modified) Demazure equations, also known as the trace constraint, $2\boldsymbol{E}_y\boldsymbol{E}_y^T\boldsymbol{E}_y - \mathrm{tr}(\boldsymbol{E}_y\boldsymbol{E}_y^T)\boldsymbol{E}_y = \boldsymbol{0}$, or explicitly,

$$e_2^2 - e_4^2 - e_5^2 + e_6^2 = 0,$$
$$e_1e_2e_6 + e_1e_4e_5 + e_3e_5^2 - e_3e_6^2 = 0, \quad (4)$$
$$e_1e_4^2 - e_1e_6^2 - e_2e_3e_6 + e_3e_4e_5 = 0,$$

as well as the rank constraint $\det(\boldsymbol{E}_y) = 0$,

$$e_1e_2e_4 + e_1e_5e_6 + e_2e_3e_5 - e_3e_4e_6 = 0. \quad (5)$$

These constraints were used in [9] to build a minimal solver for the calibrated case[2]. We also note an easy decomposition into rotation and translation components, given

---

[2]In [9] they align the z-axis with the gravity instead.

by $e_1^2 + e_3^2 = t_y^2$. In [12] it was shown that the minimal case can be solved using a single affine correspondence.

The problem becomes more difficult when adding an unknown focal length. Without any IMU data available, but still considering the partially calibrated case with only unknown focal length, the corresponding fundamental matrix has six degrees of freedom. This problem, therefore, requires a minimal case of six point correspondences, with the current state-of-the-art solver by Kukelova *et al.* [23]. Ding *et al.* [6] proposed a minimal solver for two partially calibrated cases while incorporating the IMU data. This was done by explicitly parameterizing the rotation about the gravity direction, and turning the problem into a generalized eigenvalue problem (GEP).

When assuming $\boldsymbol{R}_y = \boldsymbol{I}$, see Fig. 2, the essential matrix is $\boldsymbol{E} = [\boldsymbol{t}]_\times$, which makes the governing equations significantly easier. The minimal calibrated case requires only two point correspondences, and the epipolar constraint for a single pair of point correspondences $\boldsymbol{x} \leftrightarrow \boldsymbol{x}'$, is given by

$$\boldsymbol{x}'^T \boldsymbol{E} \boldsymbol{x} = 0 \Leftrightarrow (\boldsymbol{x} \times \boldsymbol{x}')^T \boldsymbol{t} = 0 \ . \tag{6}$$

In [10] it was also shown that the non-minimal case can be solved with global optimality guarantees.

2.2. Relative Pose with Unknown Distortion Profile

When constructing minimal solvers, it is often desirable to use as few parameters as possible. This increases robustness in RANSAC-like frameworks, as fewer iterations are needed in order to select a sample free from outliers. The one-parameter division model [7], has therefore been frequently used, as it performs well with only a single parameter for a large variety of different lenses. In this model, the radially distorted image point $\boldsymbol{x} = (x, y, 1)$ is assumed to be mapped to its corrected counterpart $\hat{\boldsymbol{x}}$ through the following parametric relation

$$\hat{\boldsymbol{x}} = f(\boldsymbol{x}, \lambda) = \begin{bmatrix} x \\ y \\ 1 + \lambda(x^2 + y^2) \end{bmatrix}, \tag{7}$$

where λ controls the level of distortion. It has been used successfully in a number of applications [22, 23, 26, 34, 35, 40].

The case of relative pose with unknown focal length and unknown distortion parameter is known to be hard. The two-sided problem, *i.e.* equal and unknown focal length and radial distortion parameter, was first studied in [14]; however, by today's standards, one cannot say that the proposed solver has much practical use: the elimination template size is very large, 886×1011, with 68 putative solutions, and a reported runtime of 400 ms. In [24] the elimination template size was reduced to 581×862 using their proposed reduction step; however, no analysis of the numerical stability was performed. Regardless, it remains impractical for

real-life applications, as the size is still exceedingly large. There has been some theoretical work on the problem, and more generally on distortion varieties [16]; however, no viable real-time solver for the case exists.

The one-sided case, *i.e.* with one calibrated camera and one camera with unknown focal length and radial distortion parameter, has been studied further. The first solver was introduced in [18], but was not numerically stable and the elimination template size was quite large, 200×231. It has later been improved in [23], and is now both numerically stable and fast, with an elimination template size of 51×70. The one-sided case, however, is mostly artificial, as it assumes one of the cameras to be calibrated, which limits the applicability of the method severely.

To the best of our knowledge, the relative pose problem with unknown and equal focal length and radial distortion parameter incorporating IMU data has not been solved. We will solve this case and show that the resulting solver is extremely fast compared to the methods discussed in this section, with an elimination template size of merely 10×21. This is done using a special assumption, which we shall discuss next.

3. Why Ignore the IMU Drift?

When measurements from the accelerometer and gyroscope are combined in an orientation filter [17, 29, 36, 36, 39] the gravity direction is preserved; however, the yaw angle begins to drift. Because of this, most visual-inertial models try to incorporate an unknown angle about the gravity direction [4–6, 9, 12, 28, 33, 38, 41]. Already in the calibrated case the governing equations (4)–(5) are non-trivial and quadratic or cubic in nature. Parameterizing the rotation matrix will also result in at least second order equations.

Instead of filtering, another approach is to simply integrate the gyroscopic data to obtain the relative orientation [8]. This eliminates error sources that potentially are introduced while fusing the gyroscopic data with the accelerometer data, *e.g.* Coriolis forces—when the IMU is not in the center of rotation—or acceleration due to motion. Both corrupt the measurement of the gravity direction. Over time integration errors, sensor bias and sensor noise will lead to drift in all axes; however, for short time intervals, this drift is very small. Additionally, if the IMU sensor has been stationary at some point in time—which is a reasonable assumption for UAVs that start from a ground position—the gyroscopic bias can be observed. The bias changes very slowly, even in low-cost consumer-available IMUs, and the observed bias can therefore be used to compensate for it in the remaining part of the sequence. In summary, we obtain very accurate relative rotations by simply integrating the gyroscopic data between camera pairs.

Valtonen Örnhag *et al.* [40] proposed using orientation filters to estimate the camera rotation and thereby bene-
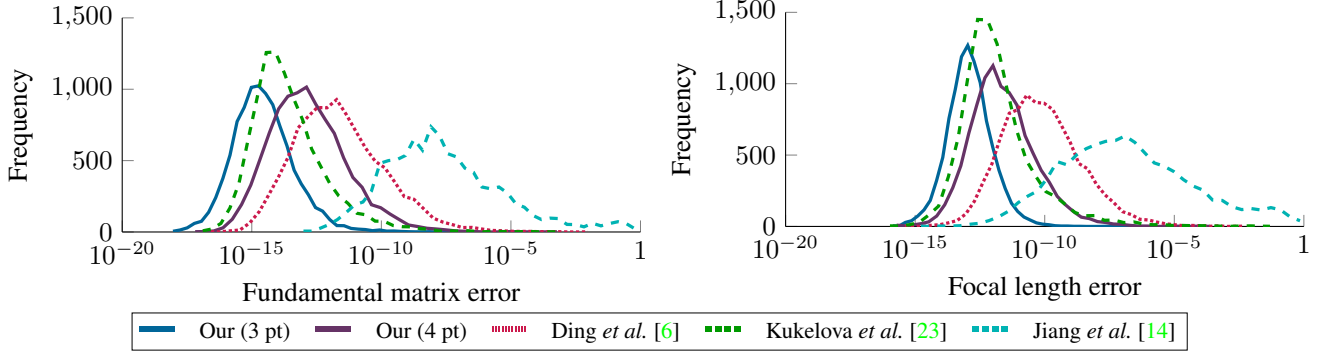
Figure 3. Error histogram for 10,000 randomly generated problem instances. The proposed 4-point method and the 7-point method [14] also solve for an unknown radial distortion parameter. The fundamental matrix error is defined as $e_{\boldsymbol{F}} := \|\boldsymbol{F} - \boldsymbol{F}_{\text{est}}\|_F / \|\boldsymbol{F}\|_F$.

fit from a relaxed problem. They showed that this allowed them to perform radial distortion correction, while still maintaining speed and accuracy. The method, however, relied on a homography-based approach, requiring planar objects in the scene geometry, thus limiting the applicability of the method.

Temporarily losing the gravity direction is not a major concern for the relative pose problem. In a complete SLAM framework, one would typically perform a visual-inertial initialization step [2, 15, 30, 32] which recovers the gravity direction as well as metric scale and bias. By trusting the IMU data we note the following:

(i) we remove one degree of freedom from the camera parameterization,

(ii) the relative pose problem becomes linear in the unknown translation, according to (6),

(iii) we open up the possibility for further intrinsic calibration and still perform fast and accurate in real-time applications.

In the next section, we will show how this is done in practice.

4. Consequences of Ignoring the IMU Drift

We construct two solvers based on our simplifying assumption. Note that the derivations are quite short, which is mainly due to the linear dependence on the translation vector.

4.1. Unknown and Equal Focal Length (3-point)

For the case of unknown and equal focal length, the fundamental matrix is given by $\boldsymbol{F} = \boldsymbol{K}^{-1}[\boldsymbol{t}]_\times \boldsymbol{R}\boldsymbol{K}^{-1}$, where $\boldsymbol{R} = \boldsymbol{R}_{\text{imu}}^{(2)}\boldsymbol{R}_{\text{imu}}^{(1)T}$ is the relative orientation. By parameterizing $\boldsymbol{K}^{-1} = \text{diag}(1,\ 1,\ w)$, we have four unknowns—the translation $\boldsymbol{t}$ and w. The epipolar constraints $\boldsymbol{x}_i^{'T}\boldsymbol{F}\boldsymbol{x}_i = 0$, for $i = 1, 2, 3$, yield three equations, which are linear in $\boldsymbol{t}$. Therefore, the resulting system can be written as

$$\boldsymbol{M}(w)\boldsymbol{t} = \boldsymbol{0}, \qquad (8)$$

where $\boldsymbol{M}(w)$ depends only on w. Even with non-degenerate configurations, the matrix $\boldsymbol{M} \in \mathbb{R}^{3\times 3}$ cannot have full rank, hence $\det(\boldsymbol{M}) = 0$. This yields a single quartic equation in the unknown w, which can be solved efficiently using the quartic root finding formula. Consequently, we have four putative solutions, from which the translation $\boldsymbol{t}$ can be retrieved by finding the null space of the 3×3 matrix $\boldsymbol{M}(w^*)$, where w^* is a putative solution. For the 3×3 case one can do this without resulting to costly SVD computations [11, 31].

4.2. Unknown and Equal Focal Length and Radial Distortion Parameter (4-point)

Assuming the one-dimensional division model (7), the (modified) epipolar constraint is given by

$$f(\boldsymbol{x}_i', \lambda)^T \boldsymbol{K}^{-1}[\boldsymbol{t}]_\times \boldsymbol{R}\boldsymbol{K}^{-1} f(\boldsymbol{x}_i, \lambda) = 0, \qquad (9)$$

for a pair of corresponding points $\boldsymbol{x}_i \leftrightarrow \boldsymbol{x}_i'$, where $\boldsymbol{K}^{-1} = \text{diag}(1,\ 1,\ w)$, with $w \neq 0$, and λ is the unknown distortion parameter. As in the previous case, we may utilize the fact that (9) is linear in $\boldsymbol{t}$, hence

$$\boldsymbol{M}(w, \lambda)\boldsymbol{t} = \boldsymbol{0}, \qquad (10)$$

where $\boldsymbol{M} \in \mathbb{R}^{4\times 3}$. This can be viewed as seeking the non-trivial nullspace of $\boldsymbol{M}$, which, in turn, implies that all 3×3 subdeterminants of $\boldsymbol{M}$ must vanish. Since there are four such subdeterminants, we can reduce the problem to four polynomial equations in two unknowns, w and λ. Furthermore, we need to exclude non-physical solutions corresponding to $w = 0$, as it turns out that there are infinitely many solutions if we allow these. This can be accomplished by saturating the corresponding ideal and can be done using the automatic Gröbner basis generator proposed in [25]. It turns out that the remaining system has eleven solutions in general; however, in practice, the most common case is that 4–6 solutions are real-valued. By using the action matrix method [3], we are able to construct an elimination template of size 10×21, by using the basis heuristic proposed in [27].
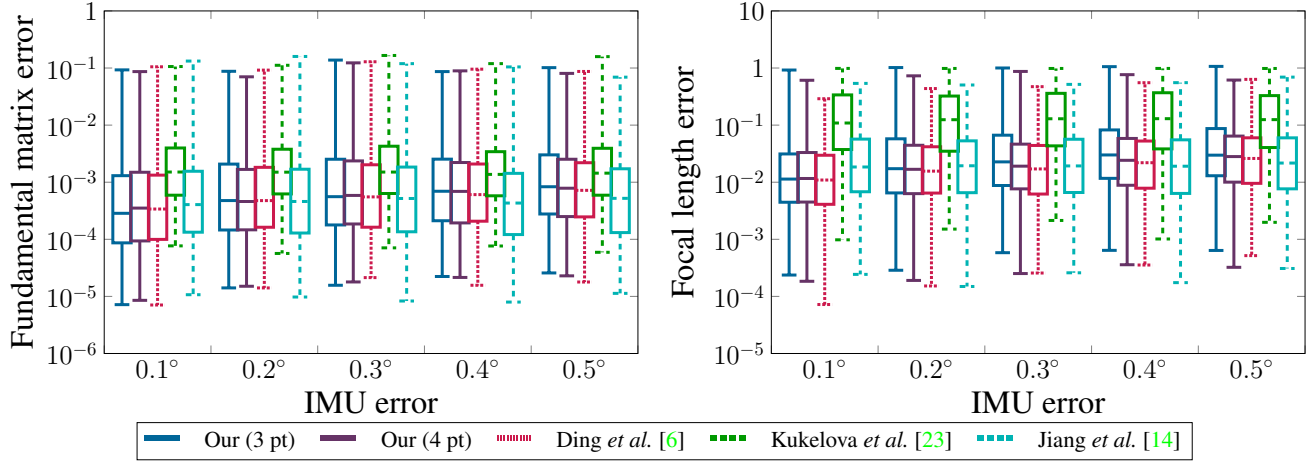
Figure 4. Error for various IMU noise levels. A total of 1,000 randomly generated problem instances are used per noise level.

5. Time Complexity

To compare timings in a fair and accurate way all solvers are implemented in C++ using the Eigen [13] library, with the same compilation flags and setup. The only exception is that the original solver by Jiang *et al.* [14] is in native MATLAB, hence significantly slower. To make a fair comparison we use the slightly faster (but numerically unstable) solver proposed in [24] which is available in C++. Note, therefore, that the real execution time for the original method would be even larger than reported. The faster solver is only used for timing, and the original solver is used for all other experiments.

To simulate a realistic environment we use a Raspberry Pi 4 to measure the execution time, as it is a fair approximation of hardware you can expect on an embedded device running these algorithms. The results are shown in Tab. 1.

Table 1. Mean execution time on a Raspberry Pi 4 for 10,000 randomly generated problem instances in C++. We also show the number of solutions for each problem. These will have to be evaluated (or at least a subset) in a RANSAC-like system, hence affect the total execution time.

Author	Time (μs)	No. Solutions
Our (3-point)	6	4
Ding *et al.* [6]	4815†	20
Kukelova *et al.* [23]	363	15
Our (4-point)	1290	11
Jiang *et al.* [14]	$1,260,700$‡	68

†: C++ implementation received from the authors of [6].

‡: Based on the smaller template reported in [24] which is numerically unstable. The original solver would be even slower.

Comparing the proposed 3-point solver to the state-of-the-art solver by Ding *et al.* [6]—essentially solving the same problem, with the exception that we ignore the po-

tential IMU drift—our solver is more than 800× faster. In addition, the proposed 4-point solver including focal length and radial distortion correction is a factor 3.7× faster than the solver by Ding *et al.* [6], and significantly faster than the solver by [14]; in fact, it is roughly 1000× faster, bridging the gap from what was considered a theoretically interesting case to something that can be applied in practice.

Let us emphasize the practical implications of simultaneously estimating the distortion parameter: the added intrinsic calibration liberates the user from time-consuming calibration procedures. This allows UAV operators (and those of other visual-inertial systems) to change optics out in the field, with no intermediate setup procedures or specific requirements needed.

6. Synthetic Experiments

In this section, we test the numerical stability and noise sensitivity of our proposed methods compared to the current state-of-the-art. The competing methods are the 4-point solver by Ding *et al.* [6], the 6-point solver by Kukelova *et al.* [23] and the 7-point solver by Jiang *et al.* [14]. We found that the solver proposed in [24], which was reported to have a smaller template size, was numerically unstable; hence we use the original method instead.

In order to get a realistic setup, random synthetic scene points with a positive depth in front of the cameras were generated. Specifically, the scene points (X, Y, Z) were uniformly distributed with $X, Z \in [-3, 3]$, the depth $Y \in [3, 8]$, and focal length $f \in [300, 3000]$. This mimics the setup used in [6]. Furthermore, the orientations are random and facing the scene. The image points are then obtained by projecting the scene points through the cameras, and the orientations of the cameras are used as input to the visual-inertial solvers. To increase numerical stability, the image point correspondences were normalized; this was

Figure 5. (*Top row*): Example images from the dataset [41]. (*Bottom row*): Images from the new dataset using the *Crazyflie 2.0*. Note that all images suffer from radial distortion to some extent.

done in the same way for all solvers. The error distribution for noise-free data is shown in Fig. 3. Here, all methods perform well, with a slight advantage for our 3-point method. Note that we include radial distortion for the proposed 4-point solver and the 7-point solver in this case as well; however, the error histograms are similar and therefore omitted.

We proceed by analyzing noise sensitivity in various situations. We test all methods on synthetic data without radial distortion (including the 4-point and 7-point solvers), and add a pixel noise relative to focal length (Gaussian noise with zero mean and standard deviation $1080/f$ pixels, where f is the ground truth focal length). Furthermore, we add noise to the IMU measurements—here we add noise on *all* angles, including the yaw angle. The noise interval is meant to cover the precision of a low-cost IMU, with a maximal error of approximately $0.5°$, as reported in [20]. The results are presented in Fig. 4. Note that both our methods perform better than the competing methods in terms of fundamental matrix recovery for smaller noise levels, and the state-of-the-art method [6] only performs slightly better for larger noise levels. This is primarily since it is capable of correcting for the error about the gravity direction. We have, however, found empirically on real-data that the lower noise levels are dominating the input data under certain conditions. This holds true even for low-cost devices, hence do not pose a practical issue.

7. Real Data

To demonstrate the applicability of our assumption, and the solvers based on it, we have used the challenging dataset [41] consisting of various indoor and outdoor scenes with predominantly planar surfaces. The scenes were cap-

tured using a mid-sized UAV (500 g, $170 \times 240 \times 40$ mm) equipped with a monochrome global shutter camera (OV9281) recorded with resolution 640×400, equipped with an onboard IMU of model MPU-9250.

In addition, we have recorded a new dataset with a different UAV. The scenes are typically demanding, *e.g.* an indoor sequence containing forward motion in a corridor, which is known to be hard. We show example images from the scenes in Fig. 5. This dataset uses a lightweight (27 g, $92 \times 92 \times 29$ mm) and low-cost nano quadcopter available under the name *Crazyflie 2.0*, captured in 640×480 resolution with an RGB camera (RunCam Nano). The ground truth was obtained using a complete SLAM system, where the reprojection error and IMU error were jointly minimized over both camera parameters and scene points, to create a globally consistent solution in metric scale. In Fig. 1 we show an image of the *Crazyflie 2.0*, as well as the output from the proposed 4-point algorithm.

The main reason to use different UAVs with different components is to show the versatility and robustness of the proposed solvers, as different setups might perform differently depending on which IMU filtering technique or pre-integration method is used. Furthermore, there are cases when the distortion profile of the camera optics may not be accurately approximated by the one-parameter division model.

In the experiments we use a simple pre-integration technique [8] to obtain the estimated relative rotation. As discussed in Sec. 3, the direction of the gravity may drift, hence the 4-point methods by Ding *et al.* [6] will not perform optimally, as one would have to select a filtering technique that preserves the gravity direction. To remedy this situation, we use the ground truth poses to align the direction for this method every frame, and apply the estimated relative ori-
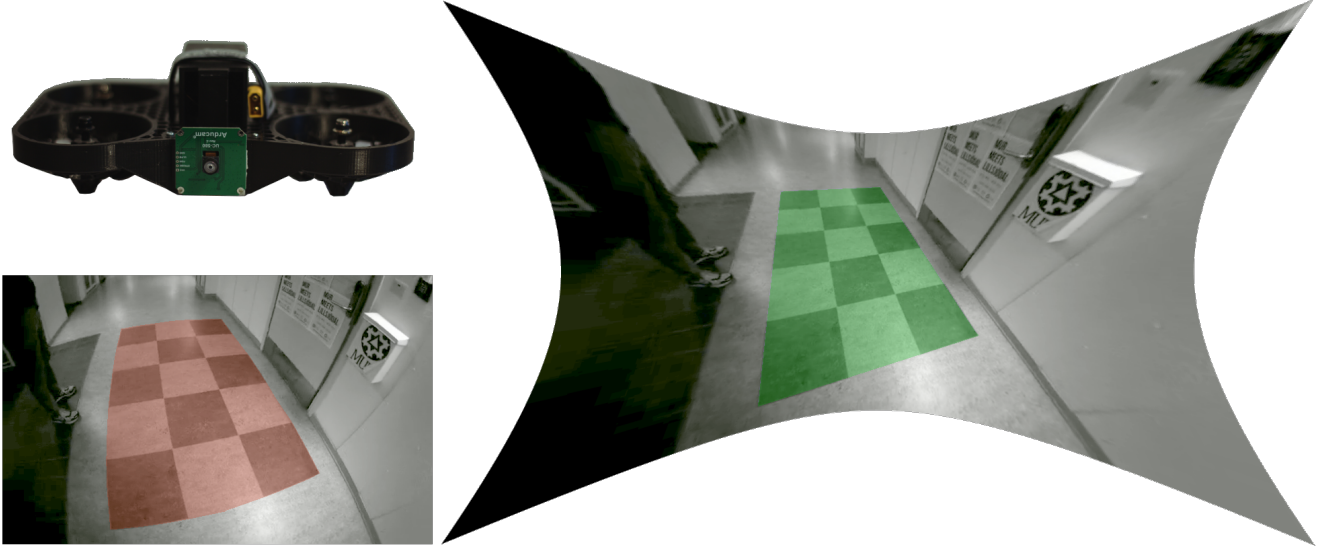
Figure 6. (*Left*:) The mid-sized UAV used in the dataset [41] and one of the input images from the *Indoor* sequence. (*Right*:) The output of the proposed 4-point algorithm, where the distortion parameter was estimated using histogram voting. Note that the checkerboard pattern on the floor is a quadrilateral box in real-life; however, it is significantly distorted (red) in the input image. The rectified image, on the other hand, displays a quadrilateral box (green). This indicates that lines are mapped to lines and that the pinhole camera model is applicable. This is strong evidence that the distortion profile has been accurately estimated, and that the one-parameter division model (7) is sufficient for the optics used on the UAV.

entation. Note that this gives the exact same relative error per input and frame as for the other solvers, but guarantees that the input to the competing 4-point solver is aligned with the gravity direction. Although these circumstances are not plausible in real-life using pre-integration, this assures that the result is not skewed by the choice of IMU filtering technique. The reported statistics for the 4-point method [6], however, are optimistic, as as the first input orientation is perfectly aligned in the experiments. In real-life situations they might perform worse.

Another important aspect to note is that we use consecutive frames, as we want to minimize potential drift. This means that the translation between frames might not be very long—typically not more than 30 cm, causing the baseline to be short. In [6] the authors only used frames $\{\{1, 11\}, \{2, 12\}, \ldots\}$ to avoid this situation; however, we want to utilize this data, as it is important for real-life applications concerning UAV positioning, *e.g.* moving obstacle avoidance.

We use the following error metrics to measure the quality of the pose reconstruction

$$e_{\boldsymbol{R}} = \arccos\left(\frac{\mathrm{tr}(\boldsymbol{R}_{\mathrm{GT}}\boldsymbol{R}_{\mathrm{est}}^T) - 1}{2}\right), \qquad (11)$$

$$e_{\boldsymbol{t}} = \arccos\left(\frac{\boldsymbol{t}_{\mathrm{GT}}^T \boldsymbol{t}_{\mathrm{est}}}{\|\boldsymbol{t}_{\mathrm{GT}}\|\|\boldsymbol{t}_{\mathrm{est}}\|}\right), \qquad (12)$$

$$e_f = \frac{|f_{\mathrm{GT}} - f_{\mathrm{est}}|}{f_{\mathrm{GT}}}, \qquad (13)$$

which have been used in a number of works [5,6,37,40,41]. The distortion profile is not as simple to measure, and we will discuss this in Sec. 7.2.

7.1. Pose Estimation with Rectified Images

In this section, we use the rectified images for the methods that do not compensate for radial distortion (the proposed 3-point solver, the 4-point solver [6] and the 6-point solver [23]), while the proposed 4-point solver is given the distorted images as input. We do not use the 7-point solver [14] in this comparison, since it is not a feasible competitor in real situations due to its computational complexity. Each method is given 1,000 RANSAC iterations with the same reprojection threshold, and no extra local optimization techniques are applied. The results are shown in Tab. 2.

From the result, we note that there is a slight advantage in terms of accuracy in favor of the method by Ding *et al.* [6] compared to the proposed 3-point solver for the dataset from [41] generated with the mid-sized UAV; however, the roles are reversed when looking at the new dataset with the *Crazyflie 2.0* UAV. Note that the rotation error for the proposed 3-point (and 4-point, since they are identical) are generally larger for the mid-sized UAV (in the range 0.4–0.6 degrees) whereas the error input to the *Crazyflie 2.0* is slightly smaller (roughly 0.2 degrees on average). This could be explained by the internal calibration between the IMU and the camera being more accurate on the *Crazyflie 2.0*. Furthermore, we would like to emphasize that the differ-

Table 2. Pose estimation error for the two datasets with rectified input images to all but the proposed 4-point method. Note that the method by [6] is given the first ground truth rotation matrix as input to ensure the assumed alignment with the y-axis. Therefore, it is likely to perform worse in real-life.

		Basement				Carpet				Indoor				Outdoor			
		Our (3 pt)	Ding	Kukelova	Our (4 pt)	Our (3 pt)	Ding	Kukelova	Our (4 pt)	Our (3 pt)	Ding	Kukelova	Our (4 pt)	Our (3 pt)	Ding	Kukelova	Our (4 pt)
Rot. Error (deg.)	Mean	0.435	**0.277**	1.336	0.435	0.628	**0.455**	5.315	0.628	0.607	**0.329**	5.582	0.607	0.423	**0.325**	4.391	0.423
	Median	0.378	**0.233**	0.765	0.378	0.566	**0.345**	2.197	0.566	0.496	**0.308**	2.169	0.496	0.345	**0.284**	1.090	0.345
Trans. Error (deg.)	Mean	5.236	**4.525**	11.649	6.127	3.296	**3.129**	11.450	4.871	4.889	**4.179**	20.648	6.093	**5.345**	7.138	25.872	6.446
	Median	3.755	**2.741**	7.457	4.458	2.243	**1.732**	8.414	2.929	2.689	**2.286**	14.811	2.722	**3.576**	4.002	17.090	4.499
Focal. Error (perc.)	Mean	24.804	**20.977**	72.814	27.315	6.921	**6.462**	63.394	46.546	**7.595**	9.011	69.189	16.190	**12.914**	27.351	88.031	39.547
	Median	6.336	**4.887**	18.681	12.293	3.597	**2.526**	23.606	7.761	4.582	**3.322**	39.160	7.546	**7.223**	8.923	39.310	9.584
		Bicycle lane				Building				Corridor				Office			
		Our (3 pt)	Ding	Kukelova	Our (4 pt)	Our (3 pt)	Ding	Kukelova	Our (4 pt)	Our (3 pt)	Ding	Kukelova	Our (4 pt)	Our (3 pt)	Ding	Kukelova	Our (4 pt)
Rot. Error (deg.)	Mean	**0.247**	1.872	3.888	0.247	**0.223**	0.861	2.016	0.223	**0.181**	0.696	1.644	0.181	**0.201**	0.871	2.935	0.201
	Median	**0.197**	0.684	1.658	0.197	**0.197**	0.413	0.902	0.197	**0.143**	0.312	0.866	0.143	**0.167**	0.576	1.398	0.167
Trans. Error (deg.)	Mean	**13.863**	23.916	27.650	19.128	**9.534**	15.077	24.944	15.181	**5.403**	10.116	17.330	7.638	**6.250**	10.091	20.422	15.616
	Median	**8.159**	13.020	21.000	10.932	**7.020**	9.296	16.750	9.526	**3.493**	5.499	11.027	4.000	**3.535**	6.465	13.490	8.326
Focal. Error (perc.)	Mean	32.138	324.201	54.166	**29.940**	**14.759**	27.875	54.659	68.120	34.890	43.671	64.270	**28.590**	**14.140**	19.288	43.134	58.680
	Median	**6.104**	10.567	22.768	10.679	**7.979**	10.456	21.067	13.702	**8.382**	10.960	27.977	11.068	**5.935**	6.024	15.369	17.952

Table 3. Pose estimation error with unrectified input images to all methods. The proposed 4-point method is the only method capable of correcting for radial distortion artifacts.

		Basement				Carpet				Indoor				Outdoor			
		Our (3 pt)	Ding	Kukelova	Our (4 pt)	Our (3 pt)	Ding	Kukelova	Our (4 pt)	Our (3 pt)	Ding	Kukelova	Our (4 pt)	Our (3 pt)	Ding	Kukelova	Our (4 pt)
Rot. Error (deg.)	Mean	**0.435**	1.117	3.750	**0.435**	**0.628**	0.989	5.445	**0.628**	**0.607**	0.798	6.555	**0.607**	**0.423**	1.042	3.546	**0.423**
	Median	**0.378**	0.669	2.782	**0.378**	**0.566**	0.729	2.828	**0.566**	**0.496**	0.542	5.428	**0.496**	**0.345**	0.635	2.507	**0.345**
Trans. Error (deg.)	Mean	25.336	29.364	20.061	**6.127**	9.248	9.437	17.567	**4.871**	18.714	20.331	24.143	**6.093**	26.498	28.659	28.739	**6.446**
	Median	16.546	22.420	14.148	**4.458**	5.443	5.049	13.328	**2.929**	9.104	11.134	18.962	**2.722**	14.735	21.101	22.681	**4.499**
Focal. Error (perc.)	Mean	2670.475	735.586	120.351	**27.315**	65.415	67.152	101.243	**46.546**	409.987	231.913	70.609	**16.190**	2659.021	718.710	123.999	**39.547**
	Median	48.684	94.448	43.793	**12.293**	62.921	56.633	50.641	**7.761**	37.991	62.136	43.290	**7.546**	60.240	129.404	48.381	**9.584**
		Bicycle lane				Building				Corridor				Office			
		Our (3 pt)	Ding	Kukelova	Our (4 pt)	Our (3 pt)	Ding	Kukelova	Our (4 pt)	Our (3 pt)	Ding	Kukelova	Our (4 pt)	Our (3 pt)	Ding	Kukelova	Our (4 pt)
Rot. Error (deg.)	Mean	**0.247**	4.175	6.883	**0.247**	**0.223**	4.263	3.852	**0.223**	**0.181**	1.674	3.221	**0.181**	**0.201**	2.377	6.975	**0.201**
	Median	**0.197**	2.205	4.280	**0.197**	**0.197**	2.137	2.738	**0.197**	**0.143**	0.795	1.809	**0.143**	**0.167**	1.065	6.048	**0.167**
Trans. Error (deg.)	Mean	45.992	42.220	38.641	**19.128**	37.619	38.630	36.473	**15.181**	17.766	23.799	24.081	**7.638**	32.616	33.743	27.831	**15.616**
	Median	47.896	37.698	32.783	**10.932**	28.936	38.025	30.757	**9.526**	10.920	15.564	17.215	**4.000**	22.109	23.946	23.512	**8.326**
Focal. Error (perc.)	Mean	1810.425	620.434	117.463	**29.940**	7189.648	455.162	80.915	**68.120**	480.445	185.039	130.248	**28.590**	4293.272	1071.105	99.617	**58.680**
	Median	250.336	204.291	44.099	**10.679**	44.812	94.533	43.344	**13.702**	39.880	58.036	51.505	**11.068**	333.428	366.350	54.794	**17.952**

ence between the proposed 3-point and competing 4-point method [6] is not very large even for the mid-sized UAV and that in a real scenario one might want to choose the proposed method, as it is $800\times$ faster. Another interesting aspect is that the visual-only method by [23] is performing significantly worse than the competing methods, which was also noted in [6].

7.2. Pose Estimation with Raw Images

We now turn our attention to using distorted input images. This scenario is interesting for UAV operators who wish to change optics out in the field without intermediate calibration procedures. The same input sequences as in Sec. 7.1 are used; however, they are not rectified prior to estimating the image point correspondences. The results are shown in Tab. 3. Unsurprisingly, our 4-point method outperforms the other methods that cannot correct for distortion artifacts. What is perhaps more interesting is that the performance, in general, is better than the visual-only 6-point method [23] on rectified data. These observations suggest that the radial distortion auto-calibration approach is practically feasible using the proposed solver.

As optics, in general, are not perfectly approximated by the one-parameter division model, it is non-trivial to express the performance of the radial distortion correction. Instead, we rely on an ocular inspection of the estimated radial distortion parameter for two sequences. In Fig. 1 and Fig. 6 we show the rectifications, using the estimated radial distortion parameter obtained from histogram voting of the respective sequence. In the latter case, we get a clear visual confirmation of the successful estimation of the radial distortion parameter, in the form of a quadrilateral checkerboard pattern visible on the floor.

8. Conclusions

In this paper, we have investigated an assumption of ignoring the IMU drift for short time intervals. We showed that modern pre-integration methods perform well and that the relative pose problem can be solved accurately and satisfactorily using this assumption. What is most important, is that the resulting equations are significantly easier to solve, opening up the possibility to tackle problems that were previously considered extremely hard and not suitable for real-time applications. We proposed the first-ever minimal solver for simultaneously estimating the focal length, distortion profile, and motion parameters while incorporating the IMU data. Furthermore, we showed a speed-up of $800\times$ compared to the current state-of-the-art for the partially calibrated case with unknown and equal focal length, with little to no loss in accuracy. The methods have been thoroughly tested on different UAVs with different components, in several challenging indoor and outdoor environments, demonstrating excellent performance.

References

[1] Snehal Bhayani, Zuzana Kukelova, and Janne Heikkila. A sparse resultant based method for efficient minimal solvers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2

[2] C. Campos, J. M. M. Montiel, and J. D. Tardós. Inertial-only optimization for visual-inertial initialization. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 51–57, 2020. 4

[3] D. A. Cox, J. Little, and D. O'Shea. *Using Algebraic Geometry*. Graduate Texts in Mathematics. Springer New York, 2005. 1, 4

[4] Yaqing Ding, Daniel Barath, and Zuzana Kukelova. Minimal solutions for panoramic stitching given gravity prior. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5579–5588, October 2021. 3

[5] Yaqing Ding, Jian Yang, Jean Ponce, and Hui Kong. An efficient solution to the homography-based relative pose problem with a common reference direction. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019. 3, 7

[6] Yaqing Ding, Jian Yang, Jean Ponce, and Hui Kong. Minimal solutions to relative pose estimation from two views sharing a common direction with unknown focal length. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2, 3, 4, 5, 6, 7, 8

[7] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, Dec 2001. 3

[8] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza. On-manifold preintegration for real-time visual–inertial odometry. *IEEE Transactions on Robotics*, 33(1):1–21, 2017. 3, 6

[9] Friedrich Fraundorfer, Petri Tanskanen, and Marc Pollefeys. A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles. In *European Conference on Computer Vision (ECCV)*, pages 269–282. Springer Berlin Heidelberg, 2010. 2, 3

[10] Johan Fredriksson, Viktor Larsson, Carl Olsson, and Fredrik Kahl. Optimal relative pose with unknown correspondences. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1728–1736, 2016. 3

[11] Theodore F. Gast, C. Fu, Chenfanfu Jiang, and J. Teran. Implicit-shifted symmetric qr singular value decomposition of 3x3 matrices. Technical report, University of California, 2016. 4

[12] Banglei Guan, Ji Zhao, Zhang Li, Fang Sun, and Friedrich Fraundorfer. Minimal solutions for relative pose with a single affine correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2, 3

[13] Gaël Guennebaud, Benoît Jacob, et al. Eigen v3. http://eigen.tuxfamily.org, 2010. 5

[14] Fangyuan Jiang, Yubin Kuang, Jan Erik Solem, and Kalle Åström. A minimal solution to relative pose with unknown focal length and radial distortion. In *Asian Conference on Computer Vision (ACCV)*, pages 443–456, 2014. 2, 3, 4, 5, 7

[15] J. Kaiser, A. Martinelli, F. Fontana, and D. Scaramuzza. Simultaneous state initialization and gyroscope bias calibration in visual inertial aided navigation. *IEEE Robotics and Automation Letters*, 2(1):18–25, 2017. 4

[16] Joe Kileel, Zuzana Kukelova, Tomas Pajdla, and Bernd Sturmfels. Distortion varieties. *Foundations of Computational Mathematics*, 18(4):1043–1071, 2018. 3

[17] E. Kraft. A quaternion-based unscented kalman filter for orientation tracking. In *Sixth International Conference of Information Fusion, 2003. Proceedings of the*, volume 1, pages 47–54, 2003. 3

[18] Y. Kuang, J. E. Solem, F. Kahl, and K. Åström. Minimal Solvers for Relative Pose with a Single Unknown Radial Distortion. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 33–40, 6 2014. 2, 3

[19] Z. Kukelova, M. Bujnak, and T. Pajdla. Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In *British Machine Vision Conference (BMVC)*, 01 2008. 2

[20] Zuzana Kukelova, Martin Bujnak, and Tomas Pajdla. Closed-Form Solutions to Minimal Absolute Pose Problems with Known Vertical Direction. In *Asian Conference on Computer Vision (ACCV)*, Queenstown, New Zealand, 11 2010. 6

[21] Z. Kukelova, M. Bujnak, and T. Pajdla. Polynomial eigenvalue solutions to minimal problems in computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1381–1393, 2012. 2

[22] Z. Kukelova, J. Heller, M. Bujnak, and T. Pajdla. Radial distortion homography. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 639–647, June 2015. 3

[23] Zuzana Kukelova, Joe Kileel, Bernd Sturmfels, and Tomás Pajdla. A clever elimination strategy for efficient minimal solvers. *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3605–3614, 2017. 2, 3, 4, 5, 7, 8

[24] V. Larsson, K. Åström, and M. Oskarsson. Efficient solvers for minimal problems by syzygy-based reduction. *Computer Vision and Pattern Recognition (CVPR)*, pages 2383–2392, July 2017. 2, 3, 5

[25] V. Larsson, K. Åström, and M. Oskarsson. Polynomial solvers for saturated ideals. *International Conference on Computer Vision (ICCV)*, pages 2307–2316, October 2017. 4

[26] Viktor Larsson, Zuzana Kukelova, and Yinqiang Zheng. Camera pose estimation with unknown principal point. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 2, 3

[27] V. Larsson, M. Oskarsson, K. Åström, A. Wallis, Z. Kukelova, and T. Pajdla. Beyond gröbner bases: Basis selection for minimal solvers. *Computer Vision and Pattern Recognition (CVPR)*, pages 3945–3954, 2018. 4

[28] B. Li, L. Heng, G. H. Lee, and M. Pollefeys. A 4-point algorithm for relative pose estimation of a calibrated camera with a known relative rotation angle. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1595–1601, 2013. 2, 3

[29] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan. Estimation of imu and marg orientation using a gradient descent algorithm. In *2011 IEEE International Conference on Rehabilitation Robotics*, pages 1–7, 2011. 3

[30] Agostino Martinelli. Closed-form solution of visual-inertial structure from motion. *International Journal of Computer Vision*, 106(2):138–152, Jan 2014. 4

[31] Aleka McAdams, A. Selle, Rasmus Tamstorf, J. Teran, and Eftychios Sifakis. Computing the singular value decomposition of 3x3 matrices with minimal branching and elementary floating point operations. Technical Report 1690, University of Wisconsin, 2011. 4

[32] R. Mur-Artal and J. D. Tardós. Visual-inertial monocular slam with map reuse. *IEEE Robotics and Automation Letters*, 2(2):796–803, 2017. 4

[33] O. Naroditsky, X. S. Zhou, J. Gallier, S. I. Roumeliotis, and K. Daniilidis. Two efficient solutions for visual odometry using directional correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):818–824, 2012. 2, 3

[34] James Pritts, Zuzana Kukelova, Viktor Larsson, and Ondrej Chum. Radially-distorted conjugate translations. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 3

[35] James Pritts, Zuzana Kukelova, Viktor Larsson, and Ondřej Chum. Rectification from radially-distorted scales. In *Asian Conference of Computer Vision (ACCV)*, pages 36–52, 2018. 3

[36] A. M. Sabatini. Quaternion-based extended kalman filter for determining orientation by inertial and magnetic sensing. *IEEE Transactions on Biomedical Engineering*, 53(7):1346–1356, 2006. 3

[37] O. Saurer, P. Vasseur, R. Boutteau, C. Demonceaux, M. Pollefeys, and F. Fraundorfer. Homography based egomotion estimation with a common direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(2):327–341, Feb 2017. 7

[38] C. Sweeney, J. Flynn, and M. Turk. Solving for relative pose with a partially known rotation is a quadratic eigenvalue problem. In *International Conference on 3D Vision (3DV)*, volume 1, pages 483–490, 2014. 2, 3

[39] Roberto Valenti, Ivan Dryanovski, and Jizhong Xiao. Keeping a good attitude: A quaternion-based orientation filter for imus and margs. *Sensors*, 15:19302–19330, 08 2015. 3

[40] Marcus Valtonen Örnhag, Patrik Persson, Mårten Wadenbäck, Kalle Åström, and Anders Heyden. Efficient real-time radial distortion correction for uavs. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1751–1760, January 2021. 3, 7

[41] Marcus Valtonen Örnhag, Patrik Persson, Mårten Wadenbäck, Kalle Åström, and Anders Heyden. Minimal solvers for indoor uav positioning. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 1136–1143, January 2021. 3, 6, 7

[42] Changchang Wu. P3.5p: Pose estimation with unknown focal length. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. 2