2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

2-Point-based Outlier Rejection for Camera-IMU Systems with applications to Micro Aerial Vehicles

Chiara Troiani[1], Agostino Martinelli[1], Christian Laugier[1] and Davide Scaramuzza[2]

***Abstract*— This paper presents a novel method to perform the outlier rejection task between two different views of a camera rigidly attached to an Inertial Measurement Unit (IMU). Only two feature correspondences and gyroscopic data from IMU measurerments are used to compute the motion hypothesis. By exploiting this 2-point motion parametrization, we propose two algorithms to remove wrong data associations in the feature-matching process for case of a 6DoF motion. We show that in the case of a monocular camera mounted on a quadrotor vehicle, motion priors from IMU can be used to discard wrong estimations in the framework of a 2-point-RANSAC based approach. The proposed methods are evaluated on both synthetic and real data.**

I. INTRODUCTION

Egomotion estimation is a core problem in the framework of many robotic applications. The estimation of the motion of a vehicle from the input of onboard cameras is known as Visual Odometry. In recent years, the coupling of vision and inertial sensing has received great attention by the mobile robotics community. It has been proved [1], [2], [3], [4], [5] that exploiting the complementarity of these sensors, it is possible to build a system capable to perform Inertial-Aided visual odometry and mapping in unknown environments.

One of the primary problems in Visual Odometry is wrong data associations. Matched features between two different camera views are usually affected by outliers. This is due to the fact that changes in viewpoint, occulsions, image noise, illumination changes and image noise are not modeled by feature-matching techiniques. To perform a robust motion estimation, it is essential to remove the outliers. The outlier detection task is usually very expensive from a computational point of view and is based on the exploitation of the geometric constraints induced by the motion model.

The standard method to process datasets contaminated by outliers is RANSAC (RANdom SAmple Consensus) [6]. It is an iterative approach consisting in the generation of hypothesis by using the minimum number of data points required to estimate the model. The generated hypothesis is then verified on the remaining subset of data and the hypothesis with the highest consensus is selected as solution.

[1]C. Troiani, A. Martinelli, and C. Laugier are with INRIA Rhone Alpes, Montbonnot, France (chiara.troiani, agostino.martinelli, christian.laugier) @inria.fr

[2]D. Scaramuzza is with the Robotics and Perception Group, University of Zurich, Switzerland—http://rpg.ifi.uzh.ch, davide.scaramuzza@ieee.org. This research was partly supported by the Swiss National Science Foundation through project number 200021-143607 (“Swarm of Flying Cameras”) and the National Centre of Competence in Research Robotics.

The number of iterations, N, necessary to guarantee that a solution free of outliers is found is [7]:

$$N = \frac{\log(1-p)}{\log(1-(1-\varepsilon)^s)} \tag{1}$$

where s is the number of data points from which the model can be computed, ε is the percentage of outliers in the dataset, p is the requested probability of success.

Table I shows the number of iterations, N, with respect to the number of points necessary to estimate the model, s. The values are computed for $p = 0.99$ and $\epsilon = 0.5$.

From Table I, we can see that the number of iterations is exponential in the minimum number of points necessary to estimate the model. For this reason, it is very important to have a minimal parametrization of the model.

TABLE I: Number of iterations of RANSAC

Number of points (s)	1	2	3	5	8
Number of iterations (N)	7	16	35	145	1177

In this paper, we presented two methods to remove outliers between two different views of a camera rigidly attached to an Inertial Measurement Unit (IMU). This paper extends our previous contribution in [8] by including the general case of a 6DoF motion. In [8] the analysis only regarded the case of a planar motion. We also show that in the case of a monocular camera mounted on a quadrotor vehicle, motion priors from IMU can be used to discard wrong estimations in the framework of a 2-point RANSAC based approach.

The paper is organized as follows. In Section II we illustrate the related works. In Section III we describe the camera motion model and the computation of the essential matrix. Section IV provides the description of the proposed algorithms for outlier rejection. Section V describes the quadrotor motion model. Section VI illustrates the performance of our methods while Section VI draws the conclusions.

II. RELATED WORKS

The minimum number of feature correspondences needed to infer a 6 degrees of freedom (6DoF) camera motion is five. This problem was solved for the first time by Kruppa [9] in 1913. Later on, many different five-point minimal solvers have been implemented [10],[11],[12], but the six [13], seven and eight-point solvers were still the most used. The more efficient five-point minimal solver was proposed by Nister [14] in 2003 and improved by Stewenius later [15]. The advantage of the five-point algorithm is that it is suitable also for planar scenarios.

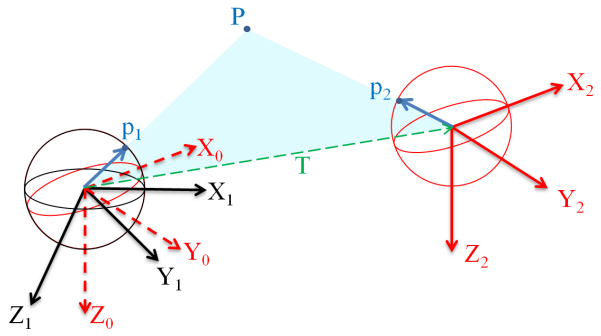

Fig. 1: Epipolar constraint. $\mathbf{p_1}$, $\mathbf{p_2}$, T and P lie on the same plane (the *epipolar plane*).

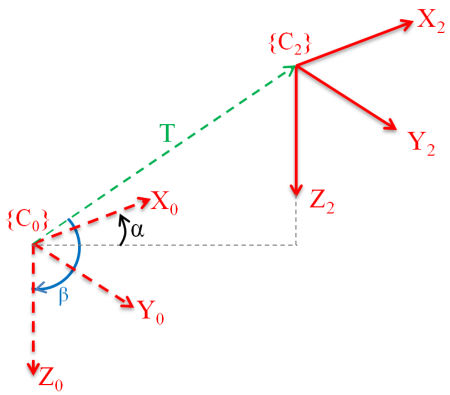

Fig. 2: The reference frame C_0 and C_2 differ only for the translation vector T. $\rho = |T|$ and the angles α and β allow us to express the origin of the reference frame C_2 in the reference frame C_0.

In the last twenty years, there has been a significant interest in minimal motion parametrizations in order to decrease the minimum number of correspondences necessary to recover the rotation and the translation (up to a scale factor) between two different camera views.

A three-point minimal solver was presented in 2010 by [16]. The authors considered a monocular camera rigidly attached to a gravity sensor. The gravity sensor can provide two out of three angles characterizing the relative orientation between two consecutive camera frames. This approach was later enhanced in [17], where the authors demonstrated the possibility of using a three-point minimal solver in a four-point RANSAC framework. The fourth point (a point distant from the scene) was used to fix the two angles of the orientation.

If the motion is constrained on a plane, its degrees of freedom decrease to three and only two points are enough to parameterize it [18]. The authors in [19], [20] demonstrated that the motion of wheeled vehicles can be considered locally circular and planar, and, consequently, characterized by two parameters. This leads to a one-point minimal solver. They proposed as well a single iteration outlier detection technique based on histogram voting.

A comparison between five, two, and one-point RANSAC algorithms for Visual Odometry can be found in [21].

III. EPIPOLAR GEOMETRY

When a camera is calibrated, it is always possible to project the feature coordinates onto a unit sphere. This allows us to make our approach independent of the camera model.

Let $\mathbf{p_1} = (x_1, y_1, z_1)$ and $\mathbf{p_2} = (x_2, y_2, z_2)$ be the image coordinates of a point feature seen from two camera positions and back projected onto the unit sphere (i.e., $\| \mathbf{p_1} \| = \| \mathbf{p_2} \| = 1$) (Figure 1).

The image coordinates of point features relative to two different unknown camera positions must satisfy the *epipolar constraint* (Figure 1) [22].

$$\mathbf{p_2}^T \mathbf{E} \mathbf{p_1} = 0 \tag{2}$$

where $\mathbf{E}$ is the *essential matrix*, defined as $\mathbf{E} = [\mathbf{T}]_\times \mathbf{R}$. $\mathbf{R}$ and $\mathbf{T} = [T_x, T_y, T_z]^T$ describe the relative rotation and translation between the two camera positions, and $[\mathbf{T}]_X$ is the skew symmetric matrix:

$$[\mathbf{T}]_\times = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \tag{3}$$

According to equation (2), the essential matrix can be computed given a set of image coordinate points. $\mathbf{E}$ can then be decomposed into $\mathbf{R}$ and $\mathbf{T}$ [22].

The minimum number of feature correspondences needed to estimate the essential matrix is function of the degrees of freedom of the camera's motion. In the case of a monocular camera performing a 6DoF motion (three for the rotation and three for the translation), considered the impossibility to recover the scale factor, a minimum of five correspondences is needed.

Let us consider a camera rigidly attached to an Inertial Measurement Unit (IMU) consisting of three orthogonal accelerometers and three orthogonal gyroscopes. The transformation between the camera reference frame $\{C\}$ and the IMU frame $\{I\}$ can be computed using [23]. Without loss of generality, we can therefore assume that these two frames are coincident ($\{I\} \equiv \{C\}$). The $\Delta\phi$, $\Delta\theta$ and $\Delta\psi$ angles characterizing the relative rotation between two consecutive camera frames can be calculated by integrating the high frequency gyroscopic measurements, provided by the IMU. This measurement is affected only by a slowly-changing drift term and can safely be recovered if the system is in motion.

If $R_x(\Delta)$, $R_y(\Delta)$, $R_z(\Delta)$ are the orthonormal rotation matrices for rotations of Δ about the x-, y- and z-axes, the matrix

$${}^{C_0}R_{C_1} = (R_x(\Delta\phi) \cdot R_y(\Delta\theta) \cdot R_z(\Delta\psi))^T \tag{4}$$

allows us to virtually rotate the first camera frame $\{C_1\}$ into a new frame $\{C_0\}$ (Figure 1) having the same orientation of the second one $\{C_2\}$.

The matrix ${}^{C_0}R_{C_1}$ allows us to express the image coordinates relative to C_1 into the new reference frame C_0:

$$\mathbf{p_0} = {}^{C_0}R_{C_1} \cdot \mathbf{p_1}. \tag{5}$$

At this point, the transformation between $\{C_0\}$ and $\{C_2\}$ is a pure translation

$$\begin{aligned}\mathbf{T} &= \rho[s(\beta)\cdot c(\alpha) \quad -s(\beta)\cdot s(\alpha) \quad c(\beta)]^T \\ \mathbf{R} &= I_3,\end{aligned} \tag{6}$$

which depends only on the angles α and β and on the scale factor ρ. The essential matrix results therefore simplified:

$$E = [\mathbf{T}]_\times \mathbf{R} = \rho \begin{bmatrix} 0 & -c(\beta) & -s(\beta)\cdot s(\alpha) \\ c(\beta) & 0 & -s(\beta)\cdot c(\alpha) \\ s(\beta)\cdot s(\alpha) & s(\beta)\cdot c(\alpha) & 0 \end{bmatrix}. \tag{7}$$

With $s(\cdot)$ and $c(\cdot)$ we denote the $\sin(\cdot)$ and $\cos(\cdot)$ respectively. At this point, being $\mathbf{p_0} = [x_0 \quad y_0 \quad z_0]^T$ and $\mathbf{p_2} = [x_2 \quad y_2 \quad z_2]^T$, the coordinates of a feature matched between two different camera frames and backprojected onto the unit sphere, we impose the epipolar constraint according to (2) and we obtain the homogeneous equation that must be satisfied by all the point correspondences.

$$\begin{aligned}x_2(y_0c(\beta) + z_0s(\alpha)s(\beta)) - y_2(x_0c(\beta) - z_0c(\alpha)s(\beta)) + \\ -z_2(y_0c(\alpha)s(\beta) + x_0s(\alpha)s(\beta)) = 0.\end{aligned} \tag{8}$$

Equation (8) depends on two parameters (α and β). This means that the relative vehicle motion can be estimated using only two image feature correspondences that we will identify as $\mathbf{p_A}$ and $\mathbf{p_B}$, where $\mathbf{p_{i_j}} = [x_{i_j} \quad y_{i_j} \quad z_{i_j}]^T$ with $i = A, B$ and $j = 0, 2$ indicate the direction of the feature i in the reference frame j.

At this point, we can recover the angles α and β solving (8) for the features $\mathbf{p_A}$ and $\mathbf{p_B}$:

$$\begin{aligned}\alpha &= -tan^{-1}\left(\frac{c_4c_2 - c_1c_5}{c_4c_3 - c_1c_6}\right), \\ \beta &= -tan^{-1}\left(\frac{c_1}{c_2c(\alpha) + c_3s(\alpha)}\right),\end{aligned} \tag{9}$$

where

$$\begin{aligned}c_1 &= x_{A_2}y_{A_0} - x_{A_0}y_{A_2}, \\ c_2 &= -y_{A_0}z_{A_2} + y_{A_2}z_{A_0}, \\ c_3 &= -x_{A_0}z_{A_2} + x_{A_2}z_{A_0}, \\ c_4 &= x_{B_2}y_{B_0} - x_{B_0}y_{B_2}, \\ c_5 &= -y_{B_0}z_{B_2} + y_{B_2}z_{B_0}, \\ c_6 &= -x_{B_0}z_{B_2} + x_{B_2}z_{B_0}.\end{aligned} \tag{10}$$

Finally, without loss of generality, we can set the scale factor ρ to 1 and estimate the essential matrix according to (7).

IV. OUTLIER REJECTION

Starting from the above-mentioned camera-motion parametrization, we propose two different approaches to reject outliers from a set of feature correspondences.

The former is based on estimating the angles α^* and β^* by selecting the pick in a Hough Space and discarding the outliers by using the reprojection error. We call this method "Hough". The second one is a 2-point RANSAC-based method with an extension to the particular case of a camera mounted on a quadrotor. The algorithm takes into account the relations between the orientation and the translation of the vehicle to remove wrong estimations. The inliers are identified using the reprojection error.

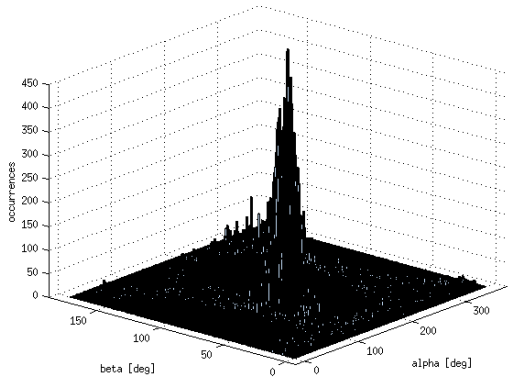

Fig. 3: Hough Space in α and β computed with real data.

A. Hough

The angles α and β are computed according to (9) from all the feature pairs matched between two consecutive frames and distant from each other more than a defined threshold (see VI). A distribution $\{\alpha_i, \beta_i\}$ with $i = 1, 2, \ldots, N$ is obtained, where N is a function of the position of the features in the environment.

To estimate the best angles α^* and β^*, we build a Hough Space (Figure 3) which bins the values of $\{\alpha_i, \beta_i\}$ into a grid of equally spaced containers. Considering that the angle β is defined in the interval $[0, \pi]$ and that the angle α is defined in the interval $[0, 2\pi]$, we set 360 bins for the variable α and 180 bins for the variable β. The number of bins of the Hough Space encodes the resolution of the estimation.

The angles α^* and β^* are therefore computed as

$$< \alpha^*, \beta^* >= \text{argmax}\{H\},$$

where H is the Hough Space.

The factors that influence the distribution are the error on the estimation of the relative rotation, the image noise, and the percentage of outliers in the data. The closer we are to ideal conditions (no noise on the IMU measurements), the narrower will be the distribution. The wider is the distribution, the more uncertain is the motion estimate.

To detect the outliers, we calculate the reprojection error relative to the estimated motion model.

The camera motion estimation can be then refined processing the remaining subset of inliers with standard algorithms [15], [22].

B. 2-point RANSAC

Using (6) we compute the motion hypothesis that consists of the translation vector $\mathbf{T}$ and the rotation matrix $\mathbf{R} = \mathbf{I}_3$ by randomly selecting two features from the correspondence set. To have a good estimation, we check that the distance between the selected features is below a defined threshold (see VI). If it is not the case, we randomly select another pair of features. Constraints on the motion of the camera can be exploited to discard wrong estimations (see Section V). The inliers are than computed using the reprojection error. The hypothesis that shows the highest consensus is considered to be the solution.

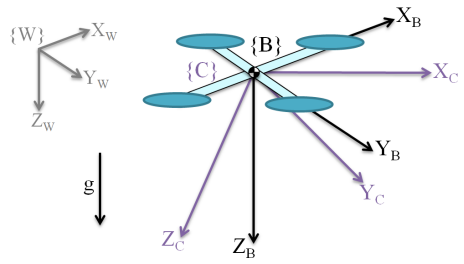

Fig. 4: Notation.

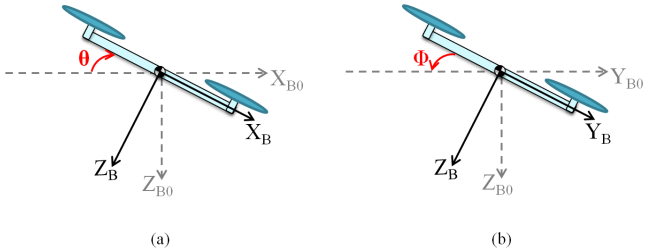

Fig. 5: Motion constraints on a quadrotor relative to its orientation. $\Delta\phi > 0$ implies a movement along Y_{B_0} positive direction, $\Delta\theta < 0$ implies a movement along Y_{B_0} positive direction.

V. QUADROTOR MOTION MODEL

We consider a quadrotor equipped with a monocular camera and an IMU.

The vehicle body-fixed coordinate frame $\{B\}$ has its Z_B-axis pointing downward (following aerospace conventions [24]). The X_B-axis defines the forward direction and the Y_B-axis follows the right-hand rule.

Without loss of generality we can consider the IMU reference frame $\{I\}$ coinciding with the vehicle body frame $\{B\}$.

The modelization of the vehicle rotation in the World frame $\{W\}$ follows the $Z-Y-X$ Euler angles convention: being ϕ, θ, ψ respectively the $Roll$, $Pitch$ and Yaw angles of the vehicle, to go from the World frame to the Body frame, we first rotate about z_W axis by the angle ψ, then rotate about the intermediate y-axis by the angle θ, and finally rotate about the X_B-axis by the angle ϕ.

The transformation between the camera reference frame $\{C\}$ and the IMU frame $\{I\}$ can be computed using [23]. Without loss of generality, we can therefore assume that also these two frames are coincident ($\{I\} \equiv \{C\} \equiv \{B\}$).

A quadrotor has 6DoF, but its translational and angular velocity are strongly coupled to its attitude due to dynamic constraints. If we consider a coordinate frame $\{B_0\}$ with the origin coincident with the one of the vehicle's body frame $\{B\}$ and the X_{B_0} and Y_{B_0} axes parallel to the ground, we observe that, in order to move in the X_{B_0} direction, the vehicle must rotate about the y-axes axis ($Pitch$ angle), while, in order to move in the Y_{B_0} direction, it must rotate about the x-axis ($Roll$ angle) (Figure 5).

These motion constraints allow us to discard wrong estimations in a RANSAC based outlier detection approach. By looking at the relation between the x and y component of the estimated translation vector and the $\Delta\phi$, $\Delta\theta$ angles provided by the IMU measurements (the same used in (4)), we are able to check the consistency of the motion hypothesis. If the estimated motion satisfies the condition

$$\begin{array}{c} ((|\Delta\phi| > \epsilon)\&(\Delta\phi \cdot T_y > 0)) \quad \| \\ ((|\Delta\theta| > \epsilon)\&(\Delta\theta \cdot T_x < 0)) \quad \| \\ ((|\Delta\phi| < \epsilon)\&(|\Delta\theta| < \epsilon)), \end{array} \tag{11}$$

we count the number of inliers (the number of correspondences that satisfy the motion hypothesis according to a predefined threshold) by using the reprojection error, otherwise we select another feature pair. The condition in (11) is satisfied if the x and y components of the motion hypothesis are coherent with the orientation of the vehicle. If both the angles $\Delta\phi$ and $\Delta\theta$ are below the threshold ϵ, we cannot infer anything about the motion and we proceed in the evaluation of the model hypothesis using the reprojection error.

The value of the threshold ϵ (see VI) is a function of the vehicle dynamics and of the controller used.

Using (1) and considering $p = 0.99$ and $\varepsilon = 0.5$, we calculate the minimum number of iterations necessary to guarantee a good performance to our algorithm and we set it to 16.

VI. RESULTS

To evaluate the performance of our algorithms, we run simulations and experiments on real data. We compared the proposed approaches with the 5-point RANSAC [14] on synthetic data, and with the 5-point RANSAC [14] and the 8-point RANSAC [25] on real data.

A. Experiments on synthetic data

We built a synthetic scenario for our simulation by using the *Robotics and Machine Vision Toolbox* for Matlab [24]. We simulated a quadrotor equipped with a downlooking monocular camera and an IMU, moving in an indoor environment (Figure 6). Random features were generated without any assumption on the structure of the environment.

The onboard downlooking monocular camera was simulated as a perspective camera with the same intrinsic parameters of the camera that we used in the experiments. A white gaussian noise with a standard deviation of 0.5 pixels was added to each extracted feature.

We generated a trajectory consisting of a take-off and of a constant-height maneuver. The camera framerate is $15Hz$, its resolution is 752 x 480. For the reprojection error in the 2-point RANSAC and in the Hough algorithm, we set a threshold of 0.5 pixels. For the 5-point RANSAC, we set the minimum number to trials to 145 iterations, and the threshold to 0.5 pixels for the reprojection error.

Figure 7 shows the results of a simulation run along the trajectory depicted in Figure 6, in the ideal case of no noisy IMU measurements. The helicopter takes off and performs a constant height maneuver.

In Figure 8, we present the results related to simulations where the quantities $\Delta\phi$, $\Delta\theta$ and $\Delta\psi$ are affected by a Gaussian noise with standard deviation of 0.3 degrees. Those errors do not affect the performance of the 5-point

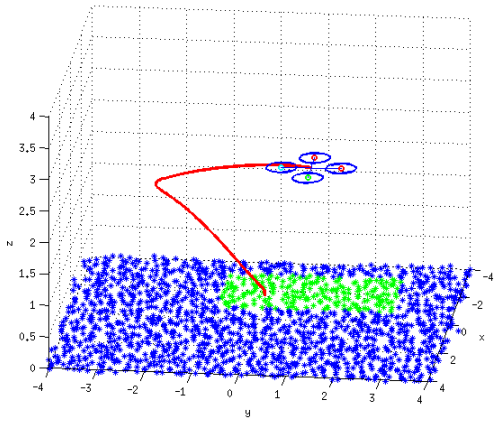

Fig. 6: Synthetic scenario. The red line represents the trajectory and the blue dots represents the simulated features. The green dots are the features in the current camera view.

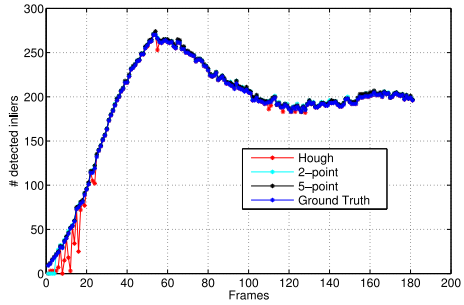

Fig. 7: The IMU measurements are not affected by noise (ideal conditions).

algorithm (that does not use IMU readings to compute the motion hypothesis). In this case, the Hough and the 2-point RANSAC approaches can still detect more than half of the inliers. The motion hypothesis can then be computed on the obtained set of correspondences by using standard approaches [15], [22].

In Figure 9, we present the results related to simulations where the quantities $\Delta\phi$ and $\Delta\theta$ are affected by a Gaussian noise with standard deviation of 0.3 degrees and in Figure 10 only the angle $\Delta\psi$ is affected by a Gaussian noise with standard deviation of 0.3 degrees. These two plots show that errors on rotations about the camera optical axis (that in our case coincides with rotations about the vehicle Z_B axis, i.e. errors on $\Delta\psi$) affects more the performances of both the algorithms than errors on $\Delta\phi$ and $\Delta\theta$.

B. Experiments on real data

The proposed approaches are tested on our nano quadrotor (Figure 11)[1] equipped with a MicroStrain 3DM-GX3 IMU ($250\ Hz$) and a Matrix Vision mvBlueFOX-MLC200w camera (FOV: $112\ deg$ and a resolution of 752 x 480).

The monocular camera calibration has been performed using the *Camera Calibration Toolbox for Matlab* [26].

To extrinsically calibrate the IMU and the camera, we used the *Inertial Measurement Unit and Camera Calibration Toolbox* [23].

[1]http://KMelRobotics.com

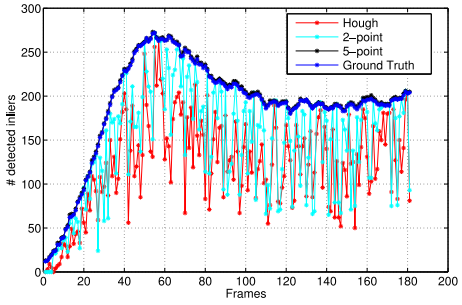

Fig. 8: The angles $\Delta\phi$, $\Delta\theta$ and $\Delta\psi$ are affected by noise.

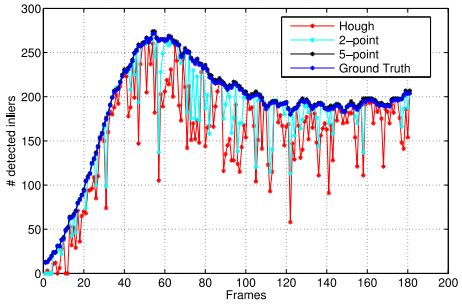

Fig. 9: Only the angles $\Delta\phi$ and $\Delta\theta$ are affected by noise.

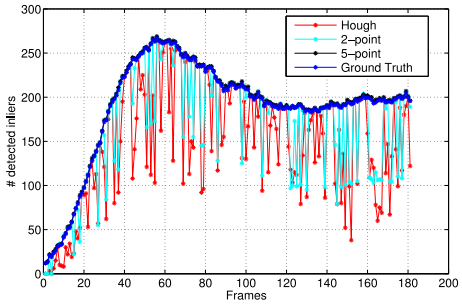

Fig. 10: Only the angle $\Delta\psi$ is affected by noise.

To validate the performance of our methods, we flew the quadrotor in our flying arena, equipped with an Optitrack motion capture system with sub-millimeter accuracy. The trajectory consisted of a take-off and a constant-height maneuver above the ground, as shown in Figure 12 and was generated using the TeleKyb Framework [27]. We recorded a dataset composed of camera images, IMU measurements and ground truth data provided by the Optitrack.

We processed our dataset with SURF features, matching them in consecutive camera frames. We run the 8-point RANSAC method on each correspondences set to have an additional term of comparison.

To evaluate the performance of our methods, we compared the number of inliers detected using the Hough and the 2-point RANSAC methods with 5-point and an 8-point RANSAC. For the 2-point RANSAC we set $\epsilon = 0.1$ deg. The results of this comparison are shown in Figure 13.

Figure 14 shows the error characterizing the estimated relative rotation between two consecutive camera frames obtained by IMU measurements and the ground truth values.

Fig. 11: Our nano quadrotor from KMelRobotics: a 150g and 18cm sized platform equipped with an integrated Gumstix Overo board and MatrixVision VGA camera.

Looking at both Figure 13 and Figure 14, we can notice that the smaller are the errors on the angles estimations, the higher is the number of inliers detected by the Hough and the 2-point RANSAC method.

Our algorithms and the algorithms that we used for the comparison, are implemented in Matlab and run on an *Intel Core i7-3740QM Processor*. We summarize their computation time in Table II.

TABLE II: Computation time

Algorithm	Hough	2-point	5-points	8-points
Time [s]	0.498	0.048	2.6869	0.0396

We can notice that the computation time of the 5-point RANSAC is almost 67 times the computation time of the 8-point RANSAC. This is due to the fact that the 5-points returns up to 10 motion solutions for each candidate set. Singular Value Decomposition (SVD) and Groebner-basis decompositions are involved and this explains the high computation time.

The computation time of the Hough algorithm is function of the number of feature pairs used to compute the distribution in Figure (3). In our experiments, we choose all the feature pairs distant more than a defined threshold one to each other. We experimentally set this threshold to 30 degrees on the unit sphere.

VII. CONCLUSIONS

In this paper, we proposed two algorithms (Hough and 2-point RANSAC) to address the outlier rejection task systems equipped with a monocular camera rigidly attached to an IMU. We used a quadorotor micro aerial vehicle as platform to demonstrate the validity of our results. We show that the relations between the vehicle's translational and angular velocity and its attitude can be exploited in order to discard wrong estimations in the framework of a RANSAC-based approach.

Both methods rely on onboard IMU measurements to calculate the relative rotation between two consecutive camera frames and to the reprojection error to detect the inliers. The two algorithms differ in the way to compute the motion hypothesis.

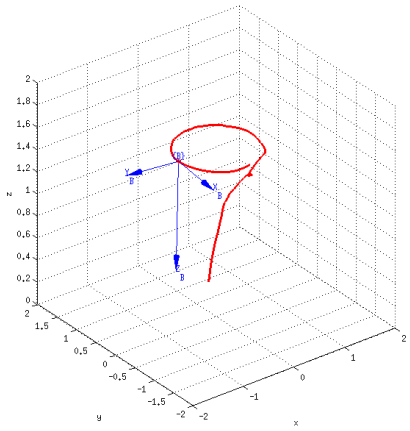

Fig. 12: Real scenario. The vehicle body frame is represented in blue, while the red line represents the followed trajectory.

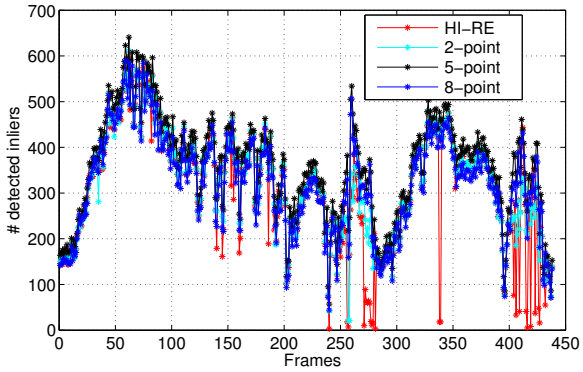

Fig. 13: Number of inliers detected with the Hough approach (red), the 2-point RANSAC (cyan), the 5-point RANSAC (black) and the 8-point RANSAC (blue) along the trajectory depicted in Figure 12.

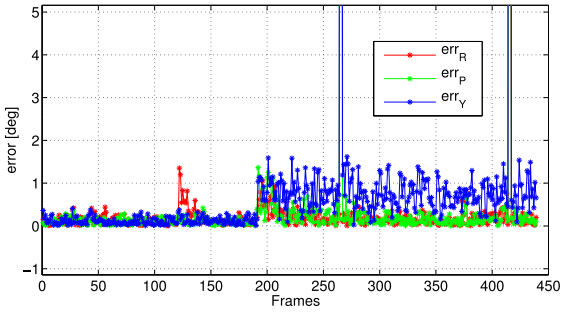

Fig. 14: Errors between the relative rotations $\Delta\phi$ (err_R), $\Delta\theta$ (err_P), $\Delta\psi$ (err_Y) estimated with the IMU and estimated with the Optitrack.

The computation time of the Hough algorithm (Table II) is function of the number of feature pairs used to compute the distribution in Figure (3). Smart policies for the choice of the pairs of features to use (based for example on the feature positions in the image plane and not only on their relative position) can be used in order to reduce the computational complexity of the approach.

Experimental results show that the 2-point RANSAC algorithm can be a good replacement of the 5-point RANSAC.

The motion hypothesis can always be refined by processing the found inliers with classic methods [15], [22].

REFERENCES

[1] S. Weiss., D. Scaramuzza, and R. Siegwart, "Monocular-slam-based navigation for autonomous micro helicopters in gps-denied environments," *Journal of Field Robotics*, vol. 28, no. 6, 2011.

[2] M. Achtelik, S. Lynen, S. Weiss, L. Kneip, M. Chli, and R. Siegwart, "Visual-inertial slam for a small helicopter in large outdoor environments," in *Video Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012.

[3] D. Scaramuzza, M. Achtelik, L. Doitsidis, F. Fraundorfer, E. Kosmatopoulos, A. Martinelli, and et al., "Vision-controlled micro flying robots: from system design to autonomous navigation and mapping in gps-denied environments, under review for the ieee robotics and automation magazine. pdf available on the author webpage." 2013.

[4] A. Martinelli, "Vision and imu data fusion: Closed-form solutions for attitude, speed, absolute scale and bias determination," *Transaction on Robotics*, vol. 28, pp. 44–60, 2012.

[5] ——, "Closed-form solution of visual-inertial structure from motion," *International Journal of Computer Vision*, vol. 106, pp. 138–152, 2014.

[6] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[7] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *Robotics & Automation Magazine, IEEE*, vol. 18, no. 4, pp. 80–92, 2011.

[8] C. Troiani, A. Martinelli, C. Laugier, and D. Scaramuzza, "1-point-based monocular motion estimation for computationally-limited micro aerial vehicles," in *ECMR*. IEEE, 2013.

[9] E. Kruppa, "Zur ermittlung eines objektes aus zwei perspektiven mit inner orientierung," in *Sitz. –Ber. Akad. Wiss, Wien, Math. Naturw. Kl., Abt. IIa.*, vol. 122, 1913, pp. 1939–1948.

[10] O. D. Faugeras and S. Maybank, "Motion from point matches: multiplicity of solutions," *International Journal of Computer Vision*, vol. 4, no. 3, pp. 225–246, 1990.

[11] J. Philip, "A non-iterative algorithm for determining all essential matrices corresponding to five point pairs," *The Photogrammetric Record*, vol. 15, no. 88, pp. 589–599, 1996.

[12] B. Triggs, "Routines for relative pose of two calibrated cameras from 5 points," 2000.

[13] O. Pizarro, R. Eustice, and H. Singh, "Relative pose estimation for instrumented, calibrated imaging platforms," in *DICTA*. Citeseer, 2003, pp. 601–612.

[14] D. Nistér, "An efficient solution to the five-point relative pose problem," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 6, pp. 756–770, 2004.

[15] H. Stewénius, C. Engels, and D. Nistér, "Recent developments on direct relative orientation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 60, no. 4, pp. 284–294, 2006.

[16] T. P. Fraundorfer F. and M. Pollefeys, "A minimal case solution to the calibrated relative pose problem for the case of two unknown orientation angles," in *European Conf. Computer Vision*, 2010, pp. 269–282.

[17] O. Naroditsky, X. S. Zhou, J. Gallier, S. I. Roumeliotis, and K. Daniilidis, "Two efficient solutions for visual odometry using directional correspondence," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 4, pp. 818–824, 2012.

[18] D. Ortin and J. Montiel, "Indoor robot motion based on monocular images," *Robotica*, vol. 19, no. 3, pp. 331–342, 2001.

[19] D. Scaramuzza, F. Fraundorfer, and R. Siegwart, "Real-time monocular visual odometry for on-road vehicles with 1-point ransac," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE, 2009, pp. 4293–4299.

[20] D. Scaramuzza, "1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints," *International journal of computer vision*, vol. 95, no. 1, pp. 74–85, 2011.

[21] ——, "Performance evaluation of 1-point-ransac visual odometry," *Journal of Field Robotics*, vol. 28, no. 5, pp. 792–811, 2011.

[22] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge Univ Press, 2000, vol. 2.

[23] J. Lobo and J. Dias, "Relative pose calibration between visual and inertial sensors," *The International Journal of Robotics Research*, vol. 26, no. 6, pp. 561–575, 2007.

[24] P. I. Corke, *Robotics, Vision & Control: Fundamental Algorithms in Matlab*. Springer, 2011.

[25] H. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms, MA Fischler and O. Firschein, eds*, pp. 61–62, 1987.

[26] J.-Y. Bouguet, "Camera calibration toolbox for matlab," 2004.

[27] V. Grabe, M. Riedel, H. Bulthoff, P. R. Giordano, and A. Franchi, "The telekyb framework for a modular and extendible ros-based quadrotor control," in *ECMR*. IEEE, 2013.