External Vision based Robust Pose Estimation System for a Quadrotor in Outdoor Environments

Wei Zheng[1], Fan Zhou[1] and Zengfu Wang[1,2]
[1]*Department of Automation, University of Science and Technology of China, Hefei, China*
[2]*Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, China*

Keywords: Vision, Pose Estimation, Quadrotor, IMU Data, Outdoors.

Abstract: In this paper, an external vision based robust pose estimation system for a quadrotor in outdoor environments has been proposed. This system could provide us with approximate ground truth of pose estimation for a quadrotor outdoors, while most of external vision based systems perform indoors. Here, we do not modify the architecture of the quadrotor or put colored blobs, LEDs on it. Only using the own features of the quadrotor, we present a novel robust pose estimation algorithm to get the accurate pose of a quadrotor. With good observed results, we get all the four rotors and calculate the pose. But when fewer than four rotors are observed, all of existing external vision based systems of the quadrotor do not mention this and could not get right pose results. In this paper, we have solved this problem and got accurate pose estimation with IMU(inertial measurement unit) data. This system can provide us with approximate ground truth outdoors. We demonstrate in real experiments that the vision-based pose estimation system for outdoor environments can perform accurately and robustly in real time.

1 INTRODUCTION

The quadrotor is a kind of micro aerial vehicle with four rotary wings. For quadrotor applications, it is necessary to set up the external pose estimation system for the quadrotor. The motivation of our work is to present an external vision based robust pose estimation system for a quadrotor in outdoor environments (see Figure 1). This system could provide us with approximate ground-truth pose of the quadrotor, which could be used to evaluate experimental results and improve algorithms. It can also be used for autonomous take-off or landing of a quadrotor. Considering occlusion or poor tracking results, most external vision based systems for the quadrotor do not mention this and can not get right pose results. We propose the novel robust pose estimation algorithms to solve this.

Nowadays, most of external vision based systems work indoors. Approaches (Altug et al., 2003; Achtelik et al., 2009; Breitenmoser et al., 2011) placed some colored blobs or LED markers on the quadrotor to estimate the pose. Colored blobs or LEDs may not be well detected and tracked in outdoor environments. (How et al., 2008; Ahrens et al., 2009; Abeywardena et al., 2013) used the VICON(a motion capture system) for observing the quadrotor indoors. These systems were complex and expensive.

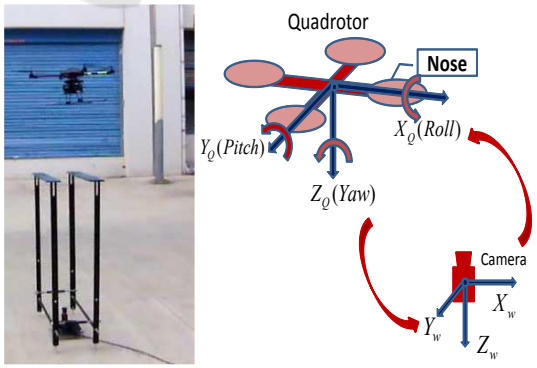


Figure 1: The robust pose estimation system for a quadrotor.

Systems designed for outdoor environments are quite rare. (Ha and Lee, 2013) used an unmanned ground vehicle to track the quadrotor. (Wendel et al., 2011; Lim et al., 2012) used results of offline SFM(structure from motion) as ground truth and could not perform in real-time.

Meanwhile, these methods (Altug et al., 2003; Achtelik et al., 2009; Breitenmoser et al., 2011) need to detect all the blobs or LEDs rightly. When one or more blobs are lost, these approaches could not get right pose estimation results.

Zheng W., Zhou F. and Wang Z..
External Vision based Robust Pose Estimation System for a Quadrotor in Outdoor Environments.
DOI: 10.5220/0004906007180723
In *Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods* (ICPRAM-2014), pages 718-723
ISBN: 978-989-758-018-5

In this paper, we detect the rotors of a quadrotor to get our reference points. This refers to the PnP(Perspective-n-Point) problem. There are some work for PnP problems (Gao et al., 2003; Dementhon and Davis, 1995; Hu and Wu, 2002; Hartley and Zisserman, 2004). The minimal numbers of correspondences to solve PnP problems is three. P4P problem with coplanar points has unique solution.

Approaches (Quan and Lan, 1999; Ansar and Daniilidis, 2003; Lepetit et al., 2008) presented linear non-iterative algorithms for PnP problems. (Lu et al., 2000; Schweighofer and Pinz, 2006) presented excellent iterative algorithms.

Recently, some significant works (Fraundorfer et al., 2010; Kukelova et al., 2010) had been done for PnP problems with IMUs.

The contribution of our work is mainly as follows. We present an external vision based robust pose estimation system for a quadrotor in outdoor environments, while most systems work indoors. Only the own features of the quadrotor are employed, while most systems modify the architecture of the quadrotor or add additional components such as LEDs. We propose the novel robust pose estimation algorithm. This algorithm could get the accurate pose from all of the four reference points with good detection or tracking results. But we may only get two or three reference points when dealing with occlusion or poor tracking results. This moment, all of existing external vision based systems for the quadrotor do not mention this and can not get right pose results. Our algorithm has solved this case and could get accurate pose estimation results by making use of the IMU data.

2 OUR WORK

2.1 The Own Features of the Quadrotor

One main character of our approach is that we only use the own features of the quadrotor. We wish that this method is simple, effective and more general. In real experiments outdoors, it is difficult to observe colored blobs, LED markers or appearance modification. The quadrotor's own features could provide us with enough information for pose estimation and are more reliable than approaches using colored markers.

2.2 Preliminary Position

Background difference and mean-shift are used for detection and tracking. A brief illustration could be seen in Figure 2. The coordinate position of the quadrotor's center is denoted by (x_q, y_q) in pixel. The

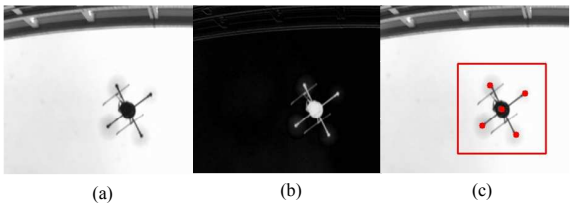


Figure 2: A brief introduction of detection and tracking.

center of CCD sensor is seen as the origin of world coordinate system. The real positions of a quadrotor are:

$$Z_q = D \cdot f / d, \tag{1}$$

$$X_q = Z_q \cdot \tilde{x}_q / f, \quad Y_q = Z_q \cdot \tilde{y}_q / f, \tag{2}$$

where f is the focal length and Z_q, X_q, Y_q are real preliminary position of the quadrotor. Here $\tilde{x}_q$ and $\tilde{y}_q$ are undistorted pixel coordinates. The D is the real wingspan of the quadrotor, and d is undistorted pixel distance of the wingspan in image. From Equation 2, some results can be obtained:

$$X_q = k \cdot Y_q \quad or \quad \tilde{x}_q = k \cdot \tilde{y}_q, \tag{3}$$

where k is scaling factor. When the intrinsic parameters of the camera are measured accurately, the scaling factor k could achieve a quite high accuracy.

2.3 Robust Pose Estimation

A simple illustration of the relative pose relationship between the camera and the quadrotor could be seen in Figure 1. Here Roll, Pitch and Yaw denote the rotation about X_Q, Y_Q and Z_Q axis.

2.3.1 Different Observed Situations

When occlusion or bad observing results occur, not all of rotors are observed. The pose estimation problems of a quadrotor could be divided into four different situations (see Figure 3).

(1): All of four rotors are observed exactly (see Figure 3 a b c).

(2): Three rotors are observed (see Figure 3 d e f).

(3): Two rotors are observed (see Figure 3 g h i).

(4): Fewer than two rotors are observed.

For these different situations, the last situation could not be solved currently. We mainly solve the former three situations. Most of current external vision based pose estimation systems of the quadrotor solve the first situation. And for the second and third situations, all of other external vision based pose estimation systems of the quadrotor do not mention this

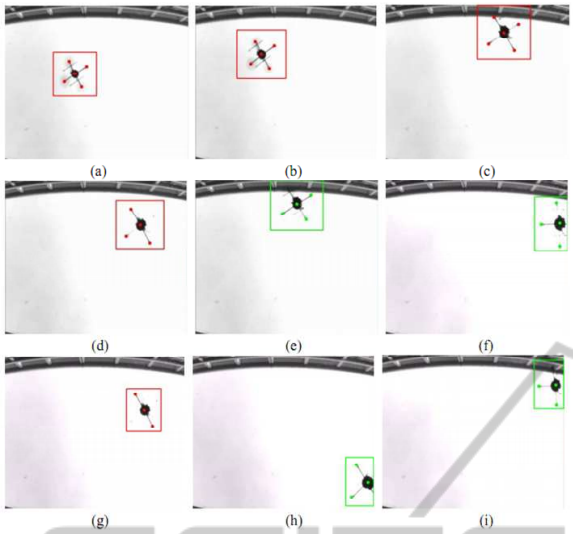


Figure 3: The different observed results of a quadrotor.

and can not get right pose results. In this paper, accurate pose estimation results could be calculated by making use of the IMU data.

2.3.2 Problem Formulation

The quadrotor could be seen as a fixed landmark when flying at different positions. Here the quadrotor is the object coordinate system. In this object coordinate system, the 3D coordinates of centers of four rotors are $M_i^o = (X_i^o, Y_i^o, Z_i^o)'$, $i = 1,2,3,4$. Centers of four rotors could be seen as coplanar points. So we could have $M_i^o = (X_i^o, Y_i^o, 0)'$, $i = 1,2,3,4$. The real values of X_i^o, Y_i^o could be measured accurately in advance. Their corresponding coordinates in camera coordinate system are $M_i^c = (X_i^c, Y_i^c, Z_i^c)'$, $i = 1,2,3,4$. We have the relationship:

$$M_i^c = R \cdot M_i^o + T, \tag{4}$$

where $R = f(\alpha, \beta, \gamma)$ and $T = (t_x, t_y, t_z)'$ are the rotation matrix and the translation vector. The centers of four rotors are expressed in the normalized image coordinate system as $m_i = (u_i, v_i, 1)'$, $i = 1,2,3,4$, which are the projection of $M_i^c = (X_i^c, Y_i^c, Z_i^c)'$, $i = 1,2,3,4$. The intrinsic parameters of the camera are known. We have relationship:

$$m_i \propto M_i^c = (R \cdot M_i^o + T). \tag{5}$$

2.3.3 EMRPP Algorithm for Four Rotors

When all of four rotors of a quadrotor are observed, the quadrotor could be seen as a landmark. The landmark formed by rotors here is four coplanar points.

(Schweighofer and Pinz, 2006) proposed a robust pose estimation algorithm(RPP) for four coplanar points. RPP made use of object-space error function. It considered that the orthogonal projection of M_i^c on m_i should be equal to M_i^c itself. This fact was shown as follows (Lu et al., 2000):

$$R \cdot M_i^o + T = F_i \cdot (R \cdot M_i^o + T), \tag{6}$$

$$F_i = m_i m_i' / m_i' m_i, \tag{7}$$

where F_i is a projection operator, $F_i = F_i'$ and $F_i = F_i^2$. The object-space error function (Lu et al., 2000) was:

$$E_{os} = \sum_{i=1}^{n} \|(I - F_i)(R \cdot M_i^o + T)\|^2. \tag{8}$$

Making use of the coplanar properties, RPP transformed the Equations 5 and 8. Then $\tilde{E}_{os}$ only depends on a rotation about the y-axis $\tilde{R}_y(\tilde{\beta})$ and $\tilde{T}$:

$$\tilde{m}_i \propto R_z(\tilde{\gamma})(\tilde{R}_y(\tilde{\beta}) \cdot \tilde{M}_i^o + \tilde{T}), \tag{9}$$

$$\tilde{E}_{os} = \sum_{i=1}^{n} ||(I - \tilde{F}_i)R_z(\tilde{\gamma})(\tilde{R}_y(\tilde{\beta}) \cdot \tilde{M}_i^o + \tilde{T})||^2, \tag{10}$$

$$\tilde{T} = (\tilde{t}_x, \tilde{t}_y, \tilde{t}_z)', \tag{11}$$

where symbol $\sim$ above the variables denotes the transformations of these variables.

Here we give the MRPP(modified RPP) algorithm. In former section, we get the preliminary position of the quadrotor. From Equation 3, we could get:

$$t_x = k \cdot t_y. \tag{12}$$

Then we put Equation 12 into Equations 5 and 8:

$$\tilde{m}_i \propto R_z(\tilde{\gamma})(\tilde{R}_y(\tilde{\beta}) \cdot \tilde{M}_i^o + \hat{T}), \tag{13}$$

$$\tilde{E}_{os} = \sum_{i=1}^{n} ||(I - \tilde{F}_i)R_z(\tilde{\gamma})(\tilde{R}_y(\tilde{\beta}) \cdot \tilde{M}_i^o + \hat{T})||^2, \tag{14}$$

$$\hat{T} = (\hat{k}\hat{t}_y, \hat{t}_y, \hat{t}_z)', \tag{15}$$

where $\hat{k}$ is the transformation of k. By using the scaling factor k, there are only two degrees of freedom in $\hat{T}$ here. Then $\tilde{E}_{os}$ only depends on a rotation about the y-axis $\tilde{R}_y(\tilde{\beta})$ and $\hat{t}_y$, $\hat{t}_z$.

As iterative algorithms, RPP and MRPP both need the initial pose guess which affects the computation time and the accuracy of results. (Lepetit et al., 2008) proposed a non-iterative solution EPnP. It had better accuracy and lower computational complexity than other non-iterative approaches.

In order to get robust and accurate pose estimation results, we present a new algorithm EMRPP combined by EPnP and our MRPP algorithms. We first perform EPnP to get the initial pose estimation. Then we make use of this initial value as the input of MRPP.

2.3.4 Three/Two Rotors Observed

When only three or two rotors of a quadrotor are observed, we will make use of the IMU data to get the accurate results. The IMU on the quadrotor could provide us with Roll and Pitch angles of the quadrotor. The angular accuracy of Roll and Pitch angles is below 0.5 degree. In general, the image data is later from -5ms to 15ms than IMU data. For pose estimation, we use the image data with the IMU data which arrived at the computer 10ms before.

Here we modify the algorithm of (Kukelova et al., 2010). In our work, the rotation matrix R is:

$$R = R_z R_y R_x, \tag{16}$$

where R_z,R_y and R_x are separately the rotation matrix for the Yaw, Pitch and Roll axis. And R_y is:

$$R_z = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{17}$$

From the data returned by the IMU, we can get the values of R_x and R_y. So the only one unknown parameter of the rotation matrix R is the rotation angle α around the Yaw axis. Equation 5 would be:

$$\lambda m_i = [R_z(\alpha) R_y R_x | T] M_i, \tag{18}$$

where λ is the scaling factor. m_i are normalized image coordinates and M_i are homogeneous coordinates. To simplify the former equation, we use the substitution $q = \tan(\alpha/2)$. Then we could get:

$$(1+q^2)R_z(q) = \begin{bmatrix} 1-q^2 & -2q & 0 \\ 2q & 1-q^2 & 0 \\ 0 & 0 & 1+q^2 \end{bmatrix}. \tag{19}$$

So the equation 18 could be written as:

$$[m_i]_\times [R_z(q) R_y R_x | T] M_i = 0, \tag{20}$$

where $[m_i]_\times$ is the skew symmetric matrix of m_i and the rank of $[m_i]_\times$ is two. Equation 20 produces three polynomial equations and only two are linearly independent. From equation 3, we have $t_x = k \cdot t_y$.

Employ this constraint in equation 20 and we get:

$$[m_i]_\times [R_z(q) R_y R_x | T(k t_y, t_y, t_z)] M_i = 0. \tag{21}$$

In this case there are only three unknown variables t_y, t_z, q. But there is variable q of degree two, the minimal number of point correspondences we need to get unique pose estimation is two.

When three rotors are observed, we have six independent polynomial equations. By using least squares method, the pose estimation could be obtained. This algorithm is named as IMU+3P. If two rotors are observed, there would be just four independent polynomial equations. So we will get unique solution of pose estimation. This algorithm is named as IMU+2P.

3 EXPERIMENTS AND RESULTS

In real experiments, we compare our algorithms with some state-of-the-art or classics algorithms, such as GAO (Gao et al., 2003), LHM (Lu et al., 2000), RPP (Schweighofer and Pinz, 2006), and EPnP (Lepetit et al., 2008).

We have performed our algorithms in real outdoor environments. Here we select the Z_q of translation results of RPP as true height. Then this true value Z_q is used in equation 2 to get the true X_q and Y_q. The real rotation angles are obtained from IMU data and electronic compass. The translation error is the angle between the estimated translation direction and the ground-truth direction. The rotation error is the smallest angle of rotation to bring the estimated rotation to the ground-truth rotation.

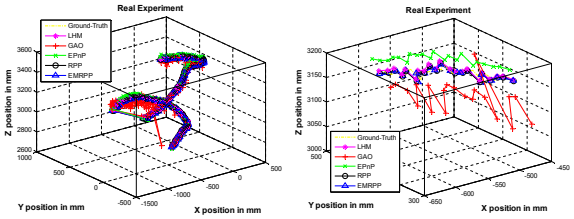
3.1 Four Rotors Observed

3.1.1 Results of Real Experiments

When all of the four rotors are observed, the results of real experiments are shows in Figure 4. The ground-truth is figured in yellow. EMRPP and RPP have better robust pose results in our real experiments. The results of RPP are close to our EMRPP. In general, EMRPP has better performance than RPP. In order to show the details of results clearly, results of 20 sequential frames are showed in Figure 4b.

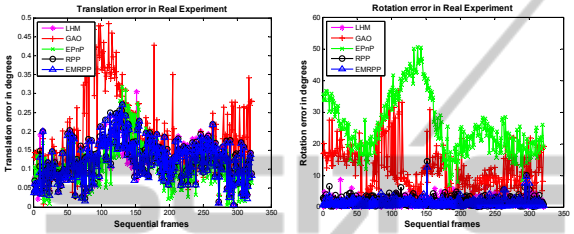
3.1.2 Pose Error in Real Experiment

Figure 5a shows the translation error in real experiments. EMRPP, LHM and RPP have better robust translation results. The translation errors of EMRPP are slightly lower than RPP. Figure 5b shows the rotation error of these algorithms in real experiments. The results of EPnP and GAO are bad. EMRPP, LHM and RPP have good rotation accuracy. EMRPP has best rotation accuracy and robust results here.



(a) (b)

Figure 4: The results of real experiments when four rotors are observed. The entire results are showed in (a), while part of results are showed in (b).



(a) (b)

Figure 5: The translation error and the rotation error.

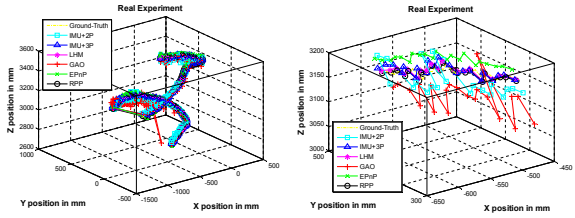
3.1.3 Discussion

Algorithms are written in C/C++ and tested in a notebook which has two i3 CPUs and 4G memory. The computation time of these algorithms can be seen in Table 1. GAO and EPnP algorithms perform fast, but they have low accuracy. EMRPP and RPP have quite good accuracy of pose estimation. Both computation time of EMRPP and RPP are below 2ms and certainly meet the request of real-time applications. In fact, our EMRPP usually has higher accuracy than RPP and performs 10% faster than RPP.

Table 1: Computation time in microsecond (μs).

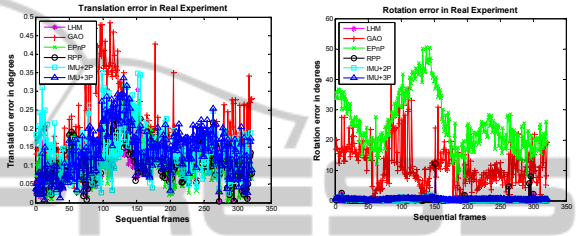
	LHM	GAO	EPnP	RPP	EMRPP
Mean	544	30	109	1739	1592
Median	508	28	99	1613	1484

Our EMRPP is combined by EPnP and MRPP. The first step of EMRPP could get accurate initial estimation of pose. A good initial estimation speeds up the following iterative section of EMRPP and get robust and accurate pose results. Because the EPnP is a fast non-iterative algorithm and the good initial estimation would speed up the MRPP algorithm. So EMRPP is faster and has higher accuracy than RPP.



(a) (b)

Figure 6: The results of real experiments when using three or two rotors. The entire results are showed in (a). Part of results are showed in (b).



(a) (b)

Figure 7: The translation error and the rotation error.

3.2 Three/Two Rotors Observed

3.2.1 Results of Real Experiments

In order to compare IMU+2P, IMU+3P with other algorithms, we also operate LHM, GAO, EPnP and RPP which use all of four point correspondences. Figure 6 shows the results of real experiments. The results of IMU+2P are better than GAO and similar to EPnP. IMU+3P, LHM and RPP have good and similar results. The accuracy of IMU+3P is quite good and meets the requirement of real applications.

3.2.2 Pose Error in Real Experiment

Figure 7a shows the translation error in real experiments. Here the translation accuracy of IMU+2P is slightly lower than LHM, RPP. IMU+3P has better translation results than IMU+2P. The accuracy of IMU+2P and IMU+3P is enough for real experiments. Figure 7b shows the rotation error in real experiment. IMU+2P and IMU+3P have the best rotation accuracy in these algorithms.

3.2.3 Discussion

IMU+3P has both good translation and rotation accuracy. IMU+2P has lower translation accuracy and higher rotation accuracy than GAO, LHM, EPnP and RPP. Usually IMU+2P has higher rotation accuracy than IMU+3P. This could be explained. The number

of point correspondences is related to translation estimation. When number of points is in a certain range, the more points we use, the higher translation accuracy we will have. IMU+3P has higher translation accuracy than IMU+2P which only applies two correspondences. And the translation error of IMU+2P and IMU+3P is usually higher than EPnP, LHM and RPP. For the rotation error, roll and pitch angles could be obtained from IMU. Usually IMU has a quite high accuracy, so IMU+2P and IMU+3P have higher rotation accuracy than GAO, LHM, EPnP and RPP. When the number of point correspondences is small and not large, the more point correspondences may disturb the accuracy of rotation calculating especially that there are high accuracy IMU data. So IMU+3P has lower rotation accuracy than IMU+2P.

4 CONCLUSIONS

In this paper, we present an external vision based robust pose estimation system for a quadrotor in outdoor environments. We only use the own features of the quadrotor. When four rotors are observed, we present the EMRPP algorithm which has higher accuracy and less computation time than RPP algorithm. When only three or two rotors are observed, we present IMU+3P or IMU+2P algorithm which could also get right pose estimation results. We have implemented real experiments using our system in outdoor environments. This system could provide us with approximate ground truth of pose for a flying quadrotor. Our pose estimation system could perform accurately and robustly in real time.

ACKNOWLEDGEMENTS

This work is supported by the National Science and Technology Major Project of the Ministry of Science and Technology of China: ITER (No.2012GB102007).

REFERENCES

Abeywardena, D., Wang, Z., Kodagoda, S., and Dissanayake, G. (2013). Visual-inertial fusion for quadrotor micro air vehicles with improved scale observability. In *ICRA*, pages 3148–3153.

Achtelik, M., Zhang, T., Kuhnlenz, K., and Buss, M. (2009). Viusal tracking and control of a quadcopter using a stereo camera system and inertial sensors. In *ICMA*, pages 2863–2869.

Ahrens, S., Levine, D., Andrews, G., and How, J. P. (2009). Vision-based guidance and control of a hovering vehicle in unknown, gps-denied environments. In *ICRA*, pages 2643–2648.

Altug, E., Ostrowski, J. P., and Taylor, C. J. (2003). Quadrotor control using dual camera visual feedback. In *ICRA*, pages 4294–4299.

Ansar, A. and Daniilidis, K. (2003). Linear pose estimation from points or lines. *In PAMI*, 25:578–589.

Breitenmoser, A., Kneip, L., and Siegwart, R. (2011). A monocular vision-based system for 6d relative robot localization. In *IROS*, pages 79–85.

Dementhon, D. F. and Davis, L. S. (1995). Model-based object pose in 25 lines of code. *In IJCV*, 15:123–141.

Fraundorfer, F., Transkanen, P., and Pollefeys, M. (2010). A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles. In *ECCV*, pages 269–282.

Gao, X.-S., Hou, X.-R., Tang, J., and Cheng, H.-F. (2003). Complete solution classification for the perspective-three-point problem. *In PAMI*, 25:930–943.

Ha, C. and Lee, D. (2013). Vision-based teleoperation of unmanned aerial and ground vehicles. In *ICRA*, pages 1465–1470.

Hartley, R. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision (Second Edition)*. Cambridge University Press.

How, J. P., Bethke, B., Frank, A., Dale, D., and Vian, J. (2008). Real-time indoor autonomous vehicle test environment. *IEEE Control Systems Magazine*, 28:51–64.

Hu, Z. and Wu, F. (2002). A note on the number of solutions of the noncoplanar p4p problem. *In PAMI*, 24:550–555.

Kukelova, Z., Bujnak, M., and Pajdla, T. (2010). Closed-form solutions to the minimal absolute pose problems with known vertical direction. In *ACCV*, pages 216–229.

Lepetit, V., Moreno-Noguer, F., and Fua, P. (2008). Epnp: Accurate non-iterative o(n) solution to the pnp problem. *In IJCV*, 81:151–166.

Lim, H., Sinha, S. N., Cohen, M. F., and Uyttendaele, M. (2012). Real-time image-based 6-dof localization in large-scale environments. In *CVPR*, pages 1043–1050.

Lu, C., Hager, G., and Mjolsness, E. (2000). Fast and globally convergent pose estimation from video images. *In PAMI*, 22:610–622.

Quan, L. and Lan, Z. (1999). Linear n-point camera pose determination. *In PAMI*, 21:774–780.

Schweighofer, G. and Pinz, A. (2006). Robust pose estimation from a planar target. *In PAMI*, 28:2024–2030.

Wendel, A., Irschara, A., and Bischof, H. (2011). Natural landmark-based monocular localization for mavs. In *ICRA*, pages 5792–5799.