A Monocular Pose Estimation Strategy for UAV Autonomous Navigation in GNSS-denied Environments

Alejandro Rodríguez-Ramos[1], Carlos Sampedro[1], Adrian Carrio[1], Hriday Bavle[1], Ramon A. Suárez Fernández[1], Zorana Milošević[1], and Pascual Campoy[1]

http://www.imavs.org/

***Abstract*—In this paper, an accurate, efficient, and simple vision-based pose estimation strategy for UAV navigation in GNSS-denied environments is presented. Using visual information and previous knowledge of 3D geometries present in the environment, the pose can be estimated accurately and used for autonomous navigation. The indoor mission in the IMAV 2016 competition has been chosen for developing and evaluating this approach. Three Perspective-n-Point (PnP) algorithms have been tested and benchmarked with the purpose of selecting the most suitable for navigating in this scenario. All of them have been tested in a realistic Gazebo-based simulation using our novel UAV software, Aerostack, which allows for a fully autonomous solution. A complete flight in a GNSS-denied environment has been successfully simulated, indicating that real flights are feasible with this approach.**

I. INTRODUCTION

Nowadays, the field of micro aerial vehicles is evolving very rapidly due to its high versatility and the capability for performing complex maneuvers in structured and unstructured environments. Even though a lot of progress has been achieved during the recent years, research efforts are still focused on obtaining fully autonomous systems that can perform high level missions such as the autonomous exploration of a GNSS-denied environment.

Performing such maneuvers in GNSS-denied environments, requires highly accurate position estimation and environment mapping. For this purpose, some approaches are based on active sensors like lasers [1]–[3], or sonars [4], [5]. However these approaches sometimes derivate in heavy weight or high power consumption, affecting agility, endurance, and range of the UAV.

For addressing this problem, some approaches for localization using vision have been proposed, using only the information provided by a single camera [6], [7]–[9], or a stereo vision system [2], [10]–[12]. These kinds of localization strategies provide the advantage of a lightweight system which can be used in indoor as well as outdoor environments. However, using a vision based system for pose estimation and navigation is a very challenging task. For this purpose, some approaches have used visual markers for retrieving the pose of the UAV [13]–[16]. In addition, optical flow based approaches have demonstrated their good capability for indoor navigation with obstacle avoidance [17]–[19].

In this paper, we present a simple and robust monocular pose estimation system for entering a building in a GNSS-denied environment by using the front camera mounted on board the UAV as the main sensor for localization. In the proposed approach, a Computer Vision algorithm has been developed for image detection of the entrances of a building, i.e. door and window, marked with colours. Once the entrances have been detected, a PnP algorithm is used for computing the pose of the UAV based on the information provided by the detection module, and the 3D geometry of door and window in a global map, which is known. We compare the localization accuracy of three different pose estimation algorithms in order to select the most suitable for an specific GNSS-denied flight scenario.

The rest of the paper is structured as follows: Section II presents the problem statement and Section III introduces the proposed methodology. In Section IV the obtained results are described. Finally, Section V concludes the paper and points towards future research directions.

II. PROBLEM STATEMENT

The strategy presented in this paper has been proposed as part of CVG-UPM team solution for participating in The International Micro Air Vehicle Conference and Competition (IMAV[1]), an international reference challenge for UAV autonomous navigation. This yearly event combines a scientific conference with a robotics competition using Micro Air Vehicles (MAVs) and focuses on research that can be applied to real life scenarios. The indoor competition for the edition of 2016 requires one or multiple UAVs to enter a building in a drilling platform to move some important objects to safe locations in a GNSS-denied environment.

In our solution for this part of the competition, we propose a monocular vision approach for pose estimation outside the building. Monocular vision can be used for pose estimation by visually recognising landmarks in the surrounding environment which can serve as navigational aids for pose estimation, providing a cheap, lightweight and low consumption solution. In this case, we propose to use the front camera mounted on

[1] All authors are with the Computer Vision Group, Center for Automation and Robotics (UPM-CSIC), Calle José Gutiérrez Abascal 2, Madrid, Spain.

[1]http://www.imavs.org

http://www.imavs.org/

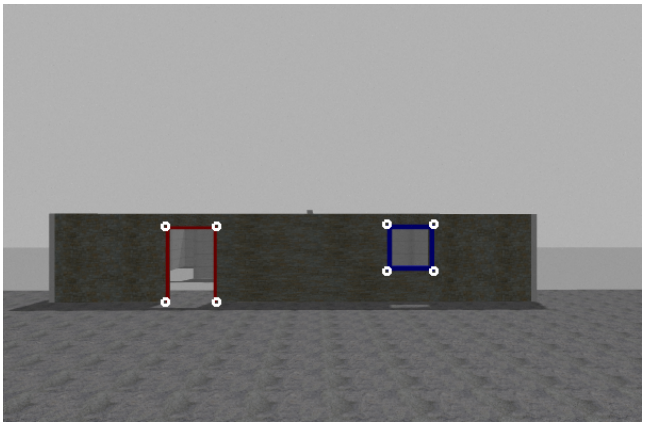


Fig. 1: Detections of door and window on the building façade.

board our UAV to capture RGB images of the building façade, detect door and window corners, whose frame of entrance will be marked with predefined colours for visual aid. The proposed detection strategy is based on two filters:

- Color filter: This filter is based on the rules proposed in the IMAV 16 competition for the entrances to the building. The door of the building has a red colored frame, while the frame of the window is blue. Taking this assumption into account the color filter of the proposed strategy is based on using the Opponent color space for enhancing the red and blue color. The proposed color channels are computed using equations 1 and 2.

$$O_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

$$O_2 = \begin{bmatrix} 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2)$$

  The next step is to apply a threshold in the corresponding color channels O_1 and O_2. Finally, an operation for finding contours is applied to the thresholded images in order to remove those contours with a small area.
- Shape filter: This filter is based on the geometry of the objects to be detected. First, an edge detector based on Canny operator is applied. After that, some morphologic operations (e.g. dilation) are applied in order to increase the detected edge contours. Finally, the geometry of the contours is taken into account for removing those which do not fit into a rectangle.

Once door and window corners have been detected, previous knowledge about their 3D geometry (in this case, their relative positions and sizes) can be used to estimate the camera pose and therefore the UAV pose, by using a Perspective-n-Point (PnP) algorithm.

The aim of the Perspective-n-Point problem is to determine the position and orientation of a camera given its intrinsic parameters and a set of n correspondences between 3D points and their 2D projections. In this case, the 3D points are the 3D coordinates of the door and window corners in an arbitrary reference system and their 2D image projections are obtained using the aforementioned detection strategy.

In the following section we present the methodology used to compare the localization accuracy given by three PnP algorithms, in an effort to determine the most suitable one for localization in this specific navigation scenario.

III. PROPOSED METHODOLOGY

In this section, the proposed PnP algorithms to be compared and the test bench structure of this analysis are presented.

A. Analyzed PnP algorithms

A set of three state-of-art PnP algorithms has been tested. As stated, these algorithms are sensitive to the number of points and the distribution of the points over the space. In the following subsection, each algorithm is presented, as well as their benefits and drawbacks in this specific scenario of study.

1) Iterative PnP: This method is based on an iterative solution based on Levenberg-Marquardt optimization. In this case, the function finds the pose of the corresponding object with respect to the camera that minimizes the re-projection error, that is, the sum of squared distances between the observed 2D projections and the projected 3D object points.

2) Robust PnP (RPnP): This algorithm is a robust non-iterative solution of PnP. This method works well for both non-redundant point sets ($n \leq 5$) and redundant point sets. In addition, this algorithm retrieves correct results robustly in the three configurations of points (Ordinary 3D case, Planar case and Quasi-singular case [20]), and its computational complexity grows linearly with n.

In our scenario of study, the points are co-planar and the number of points oscillates between 4 to 8. RPnP is capable of providing a robust and accurate solution in our case of study. Nevertheless, depending on the configuration, a non-iterative solution can lead to a lower accuracy compared to an iterative approach.

3) Efficient PnP (EPnP): [21] This method is based on a non-iterative solution to the PnP problem, in which the n 3D points of the object are expressed as a weighted sum of four virtual control points. The coordinates of these control points in the camera's reference system can be calculated by expressing these coordinates as a weighted sum of the eigenvectors of a 12×12 matrix and solving a small constant number of quadratic equations in order to obtain the right weights. This algorithm is applicable for $n \geq 4$ in planar and non-planar configurations, which make it very suitable for the problem proposed in this paper.

B. Test Bench

The test bench is based on the Aerostack architecture and a Gazebo simulated environment. All three algorithms are evaluated using a common time reference and the same set

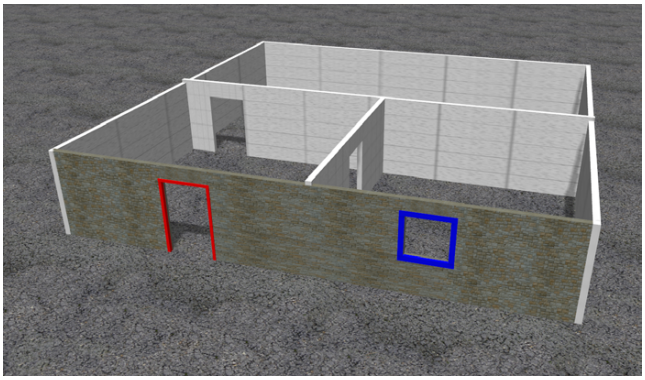


Fig. 2: IMAV'16 house replica, implemented in a Gazebo model.

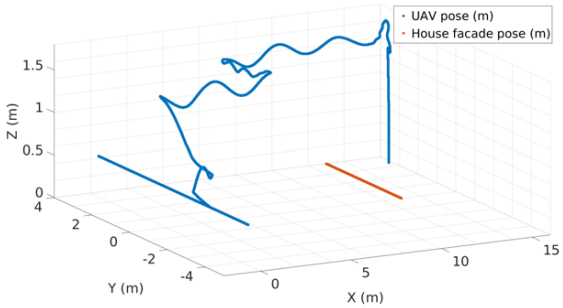


Fig. 4: UAV trajectory in the scenario of study

of 2D and 3D features. In the following subsections, the test bench is elucidated.

1) Aerostack Architecture: Aerostack (http://aerostack.org) is a software implementing fully autonomous navigation solutions for one or more heterogeneous UAVs. It is a versatile, robust and swarm-ready set of ROS packages. The autonomous navigation of the UAV in simulation has been carried out by Aerostack. A full description of Aerostack and its components falls out of the scope of this paper and can be reviewed in [22] and [23].

2) Gazebo Simulation Environment: Gazebo simulator (http://gazebosim.org) is a powerful tool to rapidly test algorithms, design robots, and perform regression testing using realistic scenarios. It provides a versatile set of models and plugins (ROS compatible), as well as an easy-to-use Application Programming Interface (API).

A replica of the IMAV'16 indoor competition set has been carefully designed in Gazebo simulator. All of the sizes of the mission elements, their colours (door and window frames) and their distribution in the scenario have been realistically modelled, in order to accurately test solutions before trying in the real world. This is best depicted in Figure 2.

Furthermore, the physical appearance and dynamics of a Parrot ARDrone 2.0 has been included, as well as different extra sensors (RGB camera, LIDAR, RGB-D camera, IMU and others). The autopilot has been simulated using a Software-In-The-Loop (SITL) simulation, in order to keep tests closer to the real world solution.

3) Simulation details: A flight from the take-off point to an area inside the house has been simulated. With a front RGB simulated camera and the simulated movements of the UAV, a realistic set of images is obtained and used to detect the corners of door and window. At any point of the flight, either the door and/or the window are seen by this camera.

In addition, Gazebo provides ground truth data of the pose of the UAV, which is used for evaluation. The trajectory followed by the UAV is shown in Figure 4.

PnP algorithms provide the pose of an arbitrary frame of reference in which the 3D points are specified with respect to the camera frame of reference. A transformation between this frame and the Gazebo global frame of reference has been taken into account for the PnP pose estimation results to be comparable with the ground truth data.

The position estimated by the three PnP algorithms and the ground truth data have been compared. Mean, standard deviation and maximum and minimum pose estimation errors have been analysed.

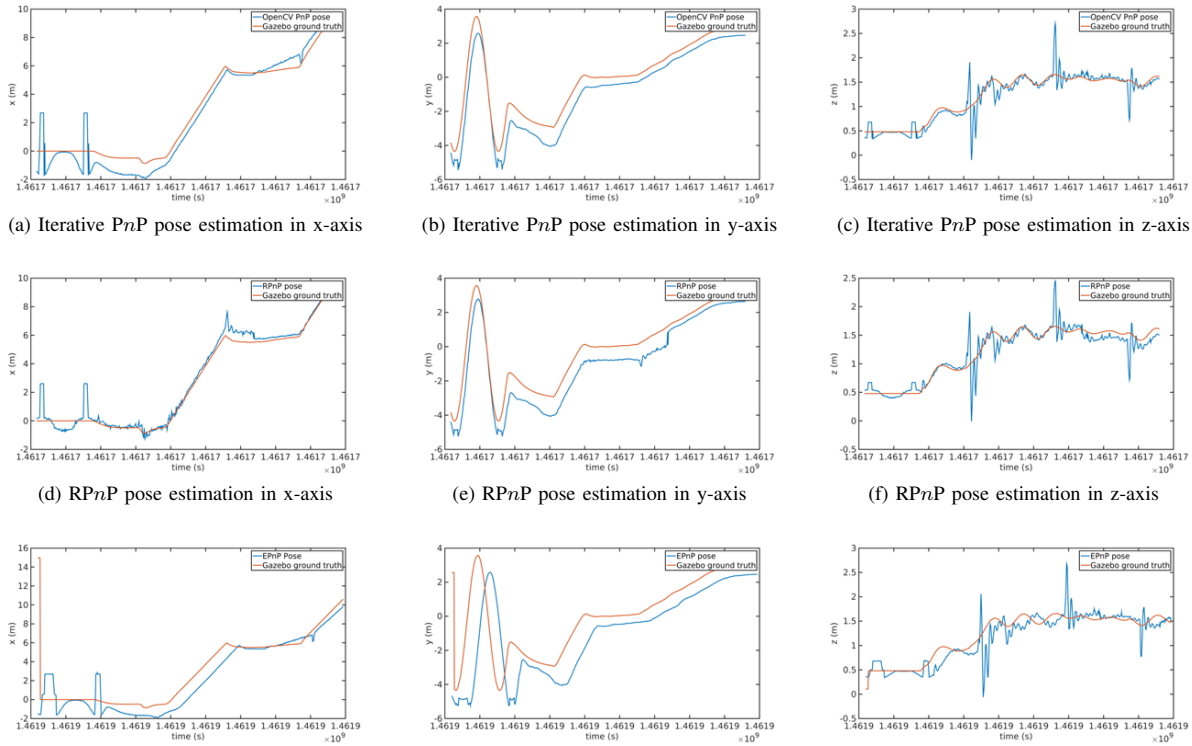
IV. RESULTS AND BENCHMARK

In this section, the benchmarking results are presented. Different error indicators have been considered in order to provide a suitable comparison among the whole set of PnP algorithms.

The difference between the ground truth data and the position estimated by the three PnP algorithms of study is best depicted in Figure 3. The whole set of algorithms shows a proper performance, regarding the ground truth data. The results are not filtered or modified by any offset, so that the PnP estimation remains unaltered. In the z-axis the results remain the closest to the ground truth. This is due to the planar configuration of the case of study, since the points share the same plane and they are perpendicular to the camera plane. This leads to an increased estimation error in x and y axes.

In addition, the absolute error for each PnP position estimation has been depicted in Figure 5. As shown, as the UAV gets closer to the house façade, the error decreases. This is due to the fact that, for a fixed camera resolution, the object size in the image increases and subsequently PnP algorithms can provide a better estimation in translation and rotation. The error in PnP algorithms is a function of the camera distance from the object space. In this planar configuration, where the points of the object space share the same plane, the error does not remain zero. PnP has some error in the estimation, due to different other sources of error, such as errors in the 2D feature detection, estimation, camera calibration, etc. Nevertheless, the error estimation is enough to properly locate the UAV in the whole environment and cross the window or the door for entering the building. This is possible by using an Extended Kalman Filter (EKF) to fuse the data of various on board

http://www.imavs.org/



(a) Iterative PnP pose estimation in x-axis (b) Iterative PnP pose estimation in y-axis (c) Iterative PnP pose estimation in z-axis

(d) RPnP pose estimation in x-axis (e) RPnP pose estimation in y-axis (f) RPnP pose estimation in z-axis

(g) EPnP pose estimation in x-axis (h) EPnP pose estimation in y-axis (i) EPnP pose estimation in z-axis

Fig. 3: Simulation results obtained in the stated experiment. (a), (b), (c), (d), (e), (f), (g), (h), (i) Pose estimation of every algorithm and Gazebo ground truth representation in every axis of the space.

sensors, such as an Inertial Measurement Unit (IMU), with the pose estimation obtained using these techniques.

Other indicators, such as Squared Root Mean Quadratic (SRMQ), maximum and minimum position estimation error can be considered. This let the benchmark to be consistent in the comparative analysis. In table I different error indicators are best enumerated.

TABLE I: Benchmark of the set of PnP algorithms.

Algorithm	SRMQ (m)	Max (m)	Min (m)	PO
OpenCV PnP	0.6571	2.696 (x-axis)	5E-4 (z-axis)	No
RPnP	0.3711	2.6081 (x-axis)	4E-5 (z-axis)	No
EPnP	2.4342	16.5752 (x-axis)	3E-05 (z-axis)	No

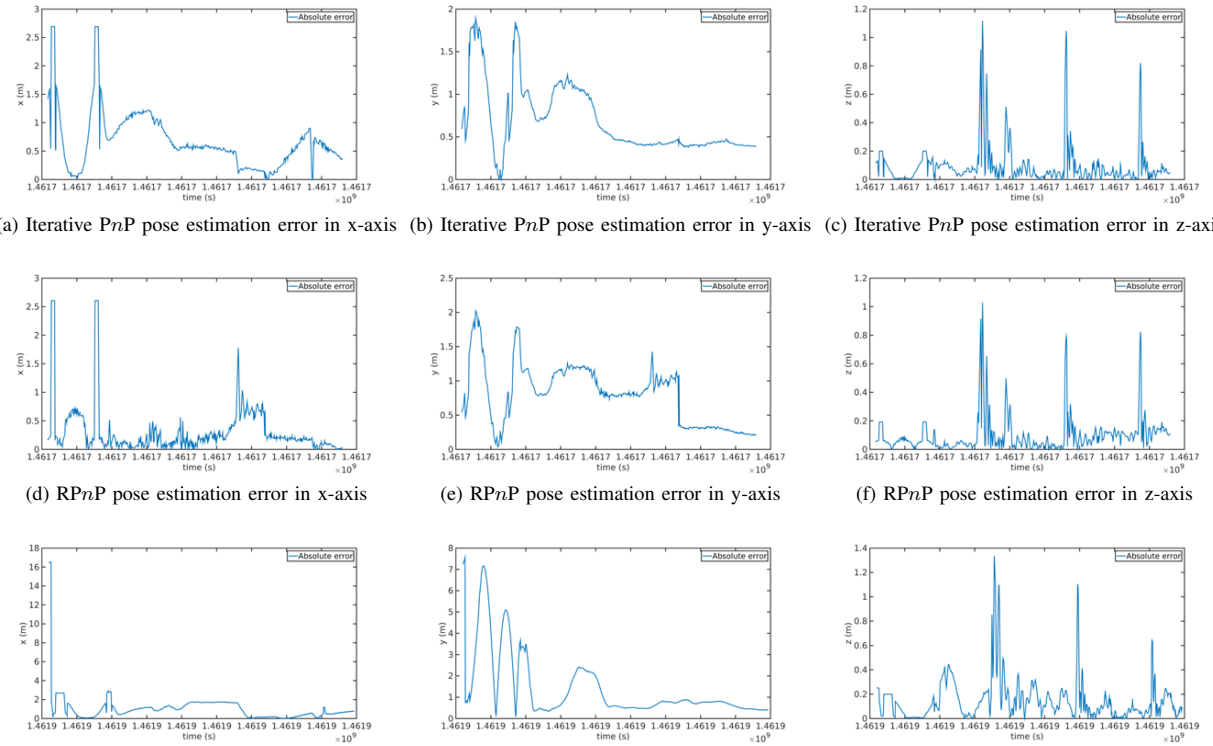
The indicators for both Iterative PnP and RPnP resulted in being really similar, although RPnP shows slightly better performance. However, the error is acceptable, especially when filtering this data and fusing it with data from other sensors. Moreover, EPnP shows poorer performance, but still valid for this application. It can be observed that none of these algorithms implement internally a prediction and optimization post process (PO). This would lead to a finer performance with more complex objects (more points spread in the object space) and does not require any other feeds of pose estimation. The RPnP algorithm performed the best in the scenario of study. Further optimizations can be made in order to decrease the stationary error.

These three state-of-art PnP algorithms have demonstrated to provide a proper performance for monocular pose estimation in GNSS-denied environments. Each of them combined with sensor fusion techniques, such as EKF, provide a complete solution for pose estimation, taking advantage of known geometries of the surrounding environment. The benchmark has revealed that the error is acceptable for real applications, with RPnP method remaining closely the best.

V. CONCLUSIONS AND FUTURE WORK

Indoor and GNSS-denied environment navigation is an extremely challenging task. To achieve this goal inexpensively without extra sensors, monocular pose estimation algorithms can be used. PnP pose estimation is based on previous knowledge of 3D geometries in the scenario for localization. Three PnP algorithms has been tested and simulated. The simulation provided proper results, showing that an autonomous entrance to an indoor environment is completely feasible with this approach. PnP has some error in the estimation, due to different other sources of error but is accurate enough to provide a precise localization at real-time frame rates. Furthermore, these results can be greatly improved when combined with other filtering and state estimation techniques.

http://www.imavs.org/



(a) Iterative PnP pose estimation error in x-axis (b) Iterative PnP pose estimation error in y-axis (c) Iterative PnP pose estimation error in z-axis

(d) RPnP pose estimation error in x-axis (e) RPnP pose estimation error in y-axis (f) RPnP pose estimation error in z-axis

(g) EPnP pose estimation error in x-axis (h) EPnP pose estimation error in y-axis (i) EPnP pose estimation error in z-axis

Fig. 5: Simulation results obtained in the stated experiment. (a), (b), (c), (d), (e), (f), (g), (h), (i) Absolute estimation error for each PnP algorithm of study.

Improvements in the detection algorithm are under study, such as making it more versatile against other types of geometries. Also, other PnP algorithms, specific for planar case, should be included, benchmarked and optimized following this work flow. Finally, real flights are required to validate our algorithm, even though results have been obtained with a highly realistic simulation.

ACKNOWLEDGMENT

This work has been developed thanks to the financial support of the Spanish Ministry of Economy and Competitiveness in the project MICYT DPI2014-60139-R.

REFERENCES

[1] R. He, S. Prentice, and N. Roy, "Planning in information space for a quadrotor helicopter in a gps-denied environment," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*. IEEE, 2008, pp. 1814–1820.

[2] M. Achtelik, A. Bachrach, R. He, S. Prentice, and N. Roy, "Stereo vision and laser odometry for autonomous helicopters in gps-denied indoor environments," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2009, pp. 733 219–733 219.

[3] S. Huh, D. H. Shim, and J. Kim, "Integrated navigation system using camera and gimbaled laser scanner for indoor and outdoor autonomous flight of uavs," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 3158–3163.

[4] J. J. Leonard and H. F. Durrant-Whyte, "Mobile robot localization by tracking geometric beacons," *Robotics and Automation, IEEE Transactions on*, vol. 7, no. 3, pp. 376–382, 1991.

[5] J. F. Roberts, T. Stirling, J.-C. Zufferey, and D. Floreano, "Quadrotor using minimal sensing for autonomous indoor flight," in *European Micro Air Vehicle Conference and Flight Competition (EMAV2007)*, no. LIS-CONF-2007-006, 2007.

[6] A. D. Wu, E. N. Johnson, M. Kaess, F. Dellaert, and G. Chowdhary, "Autonomous flight in gps-denied environments using monocular vision and inertial sensors," *Journal of Aerospace Information Systems*, 2013.

[7] G. P. Tournier, M. Valenti, J. P. How, and E. Feron, "Estimation and control of a quadrotor vehicle using monocular vision and moire patterns," in *AIAA Guidance, Navigation and Control Conference and Exhibit*, 2006, pp. 21–24.

[8] C. Bills, J. Chen, and A. Saxena, "Autonomous mav flight in indoor environments using single image perspective cues," in *Robotics and automation (ICRA), 2011 IEEE international conference on*. IEEE, 2011, pp. 5776–5783.

[9] S. Ahrens, D. Levine, G. Andrews, and J. P. How, "Vision-based guidance and control of a hovering vehicle in unknown, gps-denied environments," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE, 2009, pp. 2643–2648.

[10] C. Fu, A. Carrio, and P. Campoy, "Efficient visual odometry and mapping for unmanned aerial vehicle using arm-based stereo vision pre-processing system," in *Unmanned Aircraft Systems (ICUAS), 2015 International Conference on*. IEEE, 2015, pp. 957–962.

[11] T. Tomić, K. Schmid, P. Lutz, A. Dömel, M. Kassecker, E. Mair, I. L. Grixa, F. Ruess, M. Suppa, and D. Burschka, "Toward a fully autonomous uav: Research platform for indoor and outdoor urban search and rescue," *Robotics & Automation Magazine, IEEE*, vol. 19, no. 3, pp. 46–56, 2012.

[12] L. R. G. Carrillo, A. E. D. López, R. Lozano, and C. Pégard, "Combining

http://www.imavs.org/

stereo vision and inertial navigation system for a quad-rotor uav," *Journal of Intelligent & Robotic Systems*, vol. 65, no. 1-4, pp. 373–387, 2012.

[13] J. L. Sanchez-Lopez, J. Pestana, P. de la Puente, A. Carrio, and P. Campoy, "Visual quadrotor swarm for the imav 2013 indoor competition," in *ROBOT2013: First Iberian Robotics Conference*. Springer, 2014, pp. 55–63.

[14] J. L. Sanchez-Lopez, J. Pestana, J.-F. Collumeau, R. Suarez-Fernandez, P. Campoy, and M. Molina, "A vision based aerial robot solution for the mission 7 of the international aerial robotics competition," in *Unmanned Aircraft Systems (ICUAS), 2015 International Conference on*. IEEE, 2015, pp. 1391–1400.

[15] P. Rudol, M. Wzorek, and P. Doherty, "Vision-based pose estimation for autonomous indoor navigation of micro-scale unmanned aircraft systems," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1913–1920.

[16] A. Amor-Martinez, A. Ruiz, F. Moreno-Noguer, and A. Sanfeliu, "On-board real-time pose estimation for uavs using deformable visual contour registration," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 2595–2601.

[17] J.-C. Zufferey and D. Floreano, "Fly-inspired visual steering of an ultralight indoor aircraft," *Robotics, IEEE Transactions on*, vol. 22, no. 1, pp. 137–146, 2006.

[18] S. Zingg, D. Scaramuzza, S. Weiss, and R. Siegwart, "Mav navigation through indoor corridors using optical flow," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 3361–3368.

[19] F. Ruffier and N. Franceschini, "Visually guided micro-aerial vehicle: automatic take off, terrain following, landing and wind reaction," in *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, vol. 3. IEEE, 2004, pp. 2339–2346.

[20] S. Li, C. Xu, and M. Xie, "A robust o (n) solution to the perspective-n-point problem," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 7, pp. 1444–1450, 2012.

[21] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate o (n) solution to the pnp problem," *International journal of computer vision*, vol. 81, no. 2, pp. 155–166, 2009.

[22] J. L. Sanchez-Lopez, J. Pestana, P. de la Puente, and P. Campoy, "A reliable open-source system architecture for the fast designing and prototyping of autonomous multi-uav systems: Simulation and experimentation," *Journal of Intelligent & Robotic Systems*, pp. 1–19, 2015.

[23] J. Sanchez-Lopez, R. Suarez-Fernandez, H. Bavle, C. Sampedro, J. Pestana, M. Molina, and P. Campoy, "Aerostack: An architecture and open-source software framework for aerial robotics," in *Unmanned Aircraft Systems (ICUAS), 2016 International Conference on*, June 2016, p. 0.

http://www.imavs.org/