Towards Low-Cost Indoor Localisation Using a Multi-camera System

Oualid Araar[1(✉)], Saadi Bouhired[2], Sami Moussiou[1], and Ali Laggoune[1]

[1] Ecole Militaire Polytechnique, Bordj El Bahri, Algiers, Algeria
o.araar@emp.mdn.dz
[2] Institut de Recherche DAT, Algiers, Algeria

Abstract. Indoor localisation is a fundamental problem in robotics, which has been the subject of several research works over the last few years. Indeed, while solutions based on fusion of inertial and global navigation satellite system (GNSS) measurements have proved their efficiency in outdoor environments, indoor localisation remains an open research problem. Although commercial motion tracking systems offer very accurate position estimation, their high cost cannot be afforded by all research laboratories. This paper presents an indoor localisation solution based on a multi-camera setup. The proposed system relies on low-cost sensors, which makes it very affordable compared to commercial motion-tracking systems. We show through the experiments conducted that the proposed approach, although being cheap, can provide real-time position measurements with an error of less than 2 cm up to a distance of 2 m.

Keywords: Indoor localisation · Multi-camera system · 3D ground-truth position · Motion tracking

1 Introduction

A fundamental problem in robotics is how to obtain the position and orientation of the robot with a sufficient accuracy. Localisation solutions can be used for navigation purposes, in order to make a robot follow a desired path or trajectory. They can also be employed for obtaining the ground-truth to evaluate the accuracy of other positioning systems.

While localisation in outdoor environments has been successfully solved since GPS signal was made available to the public, solving the problem for indoor space remains a challenging task. Indeed, techniques based on the fusion of inertial and GNSS measurements have proved their efficiency for both ground and aerial vehicles navigation [1,3,5]. The poor quality or even absence of GNSS signal in indoor environments, however, makes the accuracy of these techniques very limited.

This limitation has motivated the research for other localisation techniques, which do not rely on GNSS signal. Existing solutions can be classified into

J. M. F. Rodrigues et al. (Eds.): ICCS 2019, LNCS 11540, pp. 137–148, 2019.
https://doi.org/10.1007/978-3-030-22750-0_11

two main classes, depending on whether the system is carried by the vehicle or installed externally.

While the first class offers the advantage of making the vehicle independent of any external infrastructure, it generally exhibits a drift problem due to error accumulation. The second class, on the other hand, offers drift-free measurements since no integration process is included in the position estimation. They require making modifications to the environment, which may be inconvenient in some applications.

Among commercial solutions which have been used for the purpose of indoor localisation, motion-tracking systems are the most accurate option. Such systems, which were originally developed for the purpose of tracking the joints of the human body, were quickly adopted for getting the pose of all kind of robots.

Besides their high precision which can achieve sub-millimeter accuracy, motion-tracking systems operates at high sampling rates which can attain several hundreds of FPS. Such sophisticated characteristics, however, are far from what is really needed in most robotic applications. This makes the justification of the high cost of such solutions often questionable.

The purpose of this paper is to develop a localisation approach which presents a better trade-off between performance and cost. The proposed system rely on ordinary video-cameras, which makes it very affordable.

2 Related Work

Among research works which addressed the problem of indoor position estimation using multiple cameras, the one in [11] employed a ceiling-mounted system for localising a ground robot. The overlap between pairs of cameras was exploited to get extrinsic and intrinsic parameters of each camera through stereo-calibration. A chessboard marker with side circular pattern were used for marking the robot. Given the large size and weight of the calibration chessboard, the latter cannot be carried by a robot with a limited payload. A multiple ceiling-mounted camera system was also used in [10] for tracking the position of multiple unmarked ground-robots, using a white floor to facilitate the detection of the vehicles.

Shim and Cho [12] exploited a video-surveillance system composed of multiple cameras for localising a service robot navigating in an indoor environment. In [7], a multi-agent tracking system for marked and unmarked ground-robots was presented. The cameras were mounted vertically which limits its use to only ground robots, moving on a floor perfectly parallel to the cameras plan.

Among research works which have dealt with indoor localisation of aerial vehicles the one in [6] investigated the localisation of a quadrotor UAV. This was achieved by installing a big number of markers in the laboratory and estimating the pose of the drone using an on-board camera. In [8], a multi-camera setup was used for controlling the position of a quadrotor. The vehicle was marked with four balls of different colours. In addition to the important weight of the markers, the proposed approach can only be used for localising a single agent.

Besides its low cost, the solution discussed in this paper presents the following advantages:

– Modularity: Both the hardware and software implementations are modular, which makes our solution independent of the number of sensors.
– Simplicity: The system relies on passive tags which can be easily obtained and printed on ordinary paper.
– Real-time performance: The parallel implementation of the processing algorithms allows the system to operate on full frame-rate (30 FPS).
– Opensourceness: All the third-party software used in this work is open-source, which means anyone can use it free of charge.

In the rest of this paper a detailed description of the localisation system is presented, and learned lessons from each experience are discussed. We start by a description of the different hardware components we used, followed by the optimal placement of the cameras. After that, we discuss the patterns employed for marking the object of interest as well as the algorithms allowing their detection. The estimation of the 3D position of the markers is then presented, where two techniques are implemented and compared. The last part covers the pipeline of the processing algorithms and its real-time implementation.

3 Proposed Solution

3.1 Hardware Architecture

Our localisation system uses four commercial video-cameras, Fig. 1, which provide a resolution of 640 × 480 pixels at a frame rate of 30 FPS. The cameras are interfaced to the processing laptop through a router, where each one is defined by a static IP address. Therefore, from a software point of view, adding a new sensor to the system is a matter of adding its IP address to the list of the already existing ones.

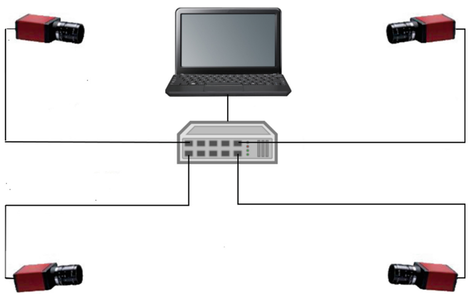


Fig. 1. Hardware components of the localisation solution.

3.2 Camera Placement

Using a multi-camera system permits enhancing the quality of the estimated position on the one hand and increasing the covered volume on the other hand. The free software "IP Video System Design Tool" was used in this work to visualise the volume covered by each sensor. In order to maximise the workspace, the sensors were placed in such a way that only two cameras cover the same space at the time. Figure 2 illustrates this configuration, where the light-green regions are covered by a single camera while at each point of the darker-green regions, the object is seen by two cameras. In this region, the position of the object can be estimated using two approaches, while in the light-green one it can only be obtained by solving the PnP problem as will be discussed in what follows.

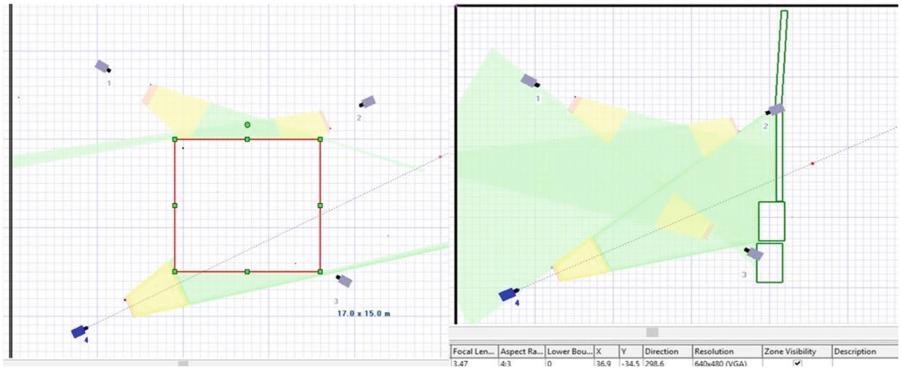
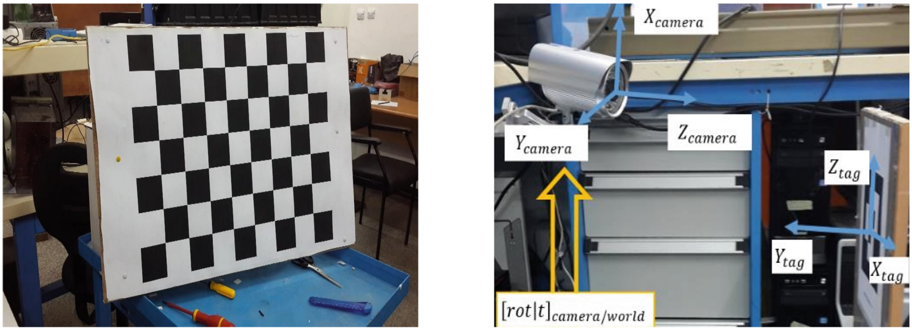


Fig. 2. Optimal placement of the four cameras obtained using the software "IP Video System Design Tool". (Color figure online)

3.3 System Calibration

Camera calibration refers to the task of determining the intrinsic parameters which appears in its projection matrix, and the extrinsic parameters which defines its pose relative to another reference frame. While intrinsic calibration was straightforward, obtaining the extrinsic parameters was a challenging task. The reason was the limited field of view between each pair of cameras, constraining the different configurations one can obtain of the calibration pattern—thing which is necessary for a precise calibration.

In order to overcome this issue we proceeded differently to the classical approach, by calibrating each camera separately, see Fig. 3. This was achieved by placing a pattern at different known poses of each camera, and then obtaining the extrinsic parameters by solving the PnP problem. Figure 4 presents a sample of calibration results using patterns of different sizes and confirms the advantages of using a big pattern.



(a) (b)

Fig. 3. System calibration: (a) stereo-calibration using a chessboard pattern; (b) calibrating each camera separately using tags.

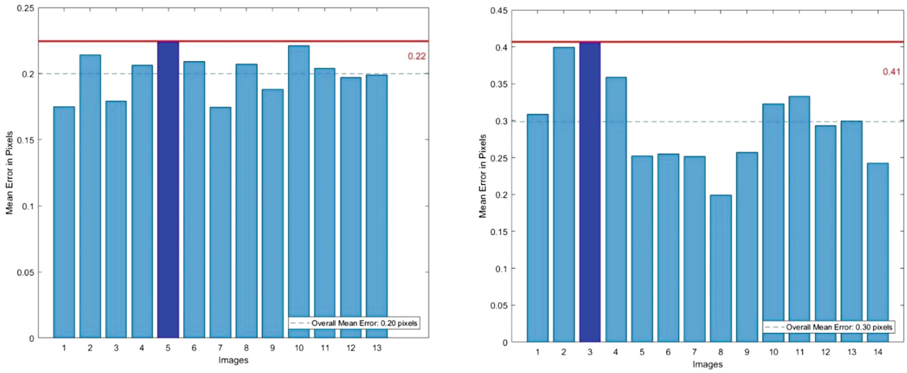
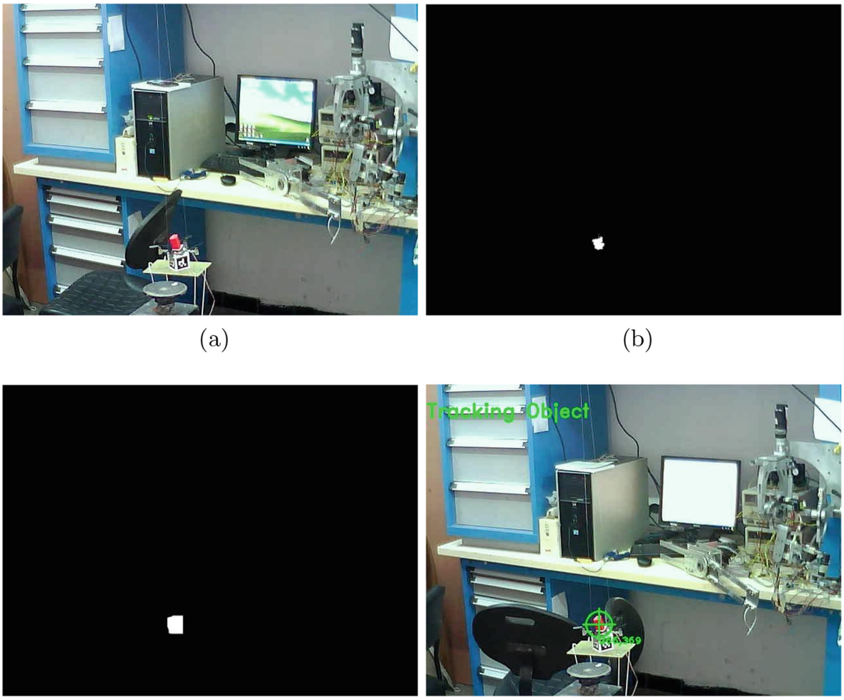


Fig. 4. Calibration results obtained using different pattern sizes. (left) big pattern, (right) smaller one.

3.4 Object Detection and Identification

In order to facilitate the detection of an object of interest and determine its identity in case of multi-object application, two kinds of artificial markers were employed. The first one is a coloured marker while the second is a tag from the AprilTag library. In what follows, we discuss the detection and identification of each of the two markers.

Coloured Marker. For detecting the coloured marker, the images are first transformed from the RGB to the HSV space where the detection thresholds are specified by the user. Pixels within these thresholds represent eventual candidates and are thus retained for the next operation (Fig. 5b). After a dilation/erosion step only regions with the colour of interest are conserved (Fig. 5c). The centroid of the object of interest is calculated based on image-moment, Fig. 5d.



(a) (b) (c) (d)

Fig. 5. Coloured marker detection steps: (a) raw images; (b) detection in HSV space; (c) noise suppression after erosion/dilation step; (d) centroid calculation.

Tag Markers. Among existing visual tag systems, one can mention the ARToolkit [4], Studier-stube Tracker [13], ARTag [2], and the more recent AprilTag system [9]. The latter is chosen in this work, as it enables encoding a large number of distinct ids. Furthermore, its encoding scheme enables codewords with a low hamming distance, which results in a lower misidentification rate.

The AprilTag system works in two steps: detection and decoding. The detector begins by clustering pixels based on their gradient magnitude and direction, to detect line segments. Sequences of lines that form a 4-sided shape (quad) are then identified, see Fig. 6.

The decoding stage reads the bits from each tag's payload field, in order to identify it. This stage begins by transforming each bit field into image coordinates using the Homography matrix estimated from the four corners of the detected quad. The resulting pixels are then classified using a spatially varying model of the intensities of the white and black colours.

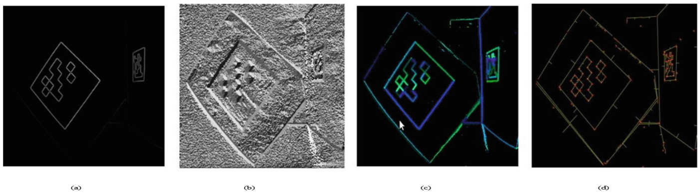


Fig. 6. Detection steps of the tags from the AprilTag library.

3.5 Position Estimation

The detection and identification steps allow obtaining the 2D coordinates of the object of interest in the image space. In order to get the 3D position of the object relative to the scene, two approaches can be considered. In the first, measurements from at least two cameras are combined in order to get the depth of the object, using what is commonly referred to as triangulation. In the second approach, the position is estimated from each camera separately by solving what is known as the perspective n point (PnP) problem. Both approaches are detailed in what follows.

Triangulation. Unlike RGBD cameras which provides the full 3D coordinates of a point of interest, only the direction of the object can be obtained from a video-camera. If measurements of the same object are available from a second sensor, its 3D position can be obtained from the intersection of the two lines of sight, Fig. 7. This approach is employed for estimating the coloured marker position, since just the 2D coordinates of its centroid are available and thus its 3D position can only be estimated using at least two cameras.

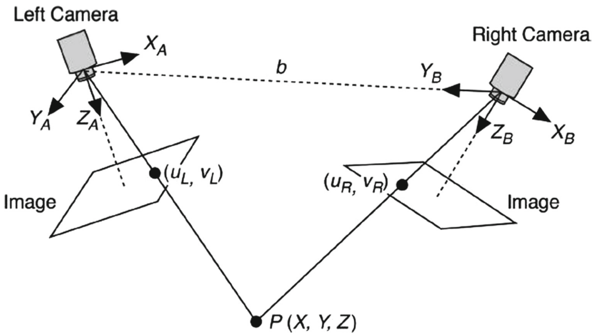


Fig. 7. Object position estimation using Triangulation.

Solving PnP Problem. Unlike the coloured marker, the identified tags provide the 2D coordinates of four points from each tag. Given the prior knowledge about the dimension and shape of the tags, one can estimate the pose of the marked object from a single camera. This problem, which is commonly coined PnP, can be solved if a minimum of four points is available from the object of interest.

3.6 Real-Time Implementation

In order to ensure real-time performance, all the previous steps where implemented in C++, based on the Open Computer Vision (OpenCV) library. A serial implementation of the previously discussed processing algorithms turned-out to be impractical. In fact, even without any processing, i.e only the acquisition and display of the images, an important lag among the four cameras was observed. To overcome this issue, we proceeded to a parallel implementation, using the Microsoft TBB library. An overview of the architecture of this implementation is given in Fig. 8. Using this approach, we were able to run the code in full frame rate on a laptop with an $i7$ 2.5 GHz processor and 8 GB of RAM.

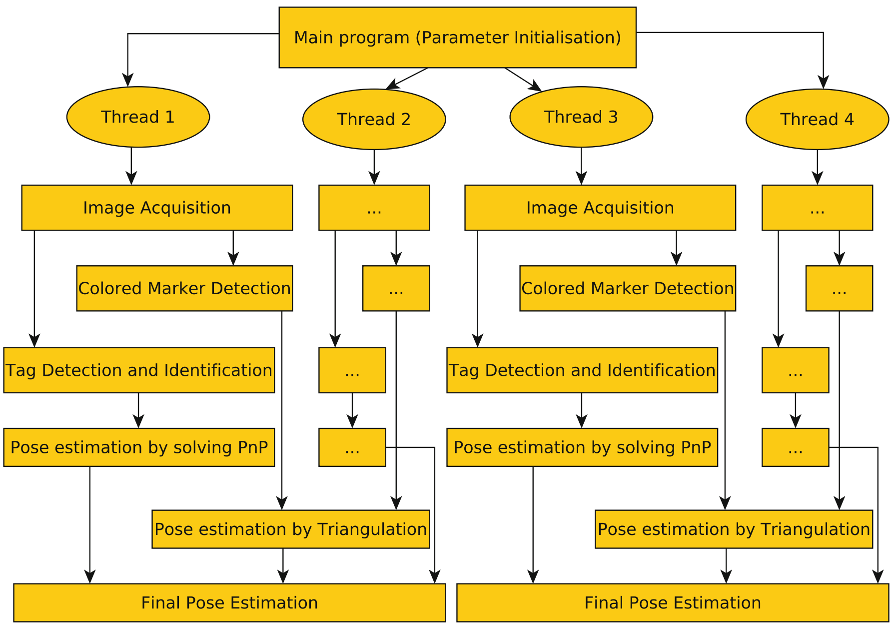


Fig. 8. Architecture of the real-time implementation of the processing algorithms.

4 Experimental Results

In order to validate the system, the experimental set-up of Fig. 9 was realised. The four cameras were installed in the optimal configuration discussed previously.

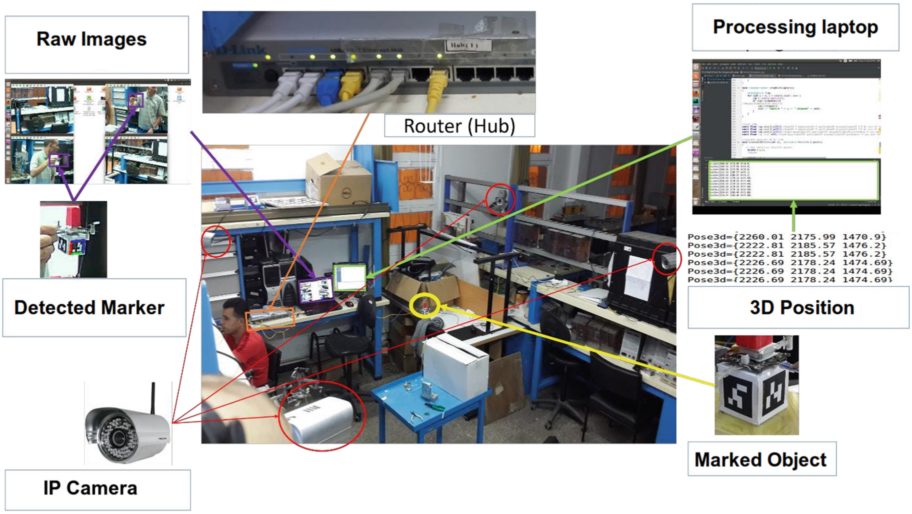


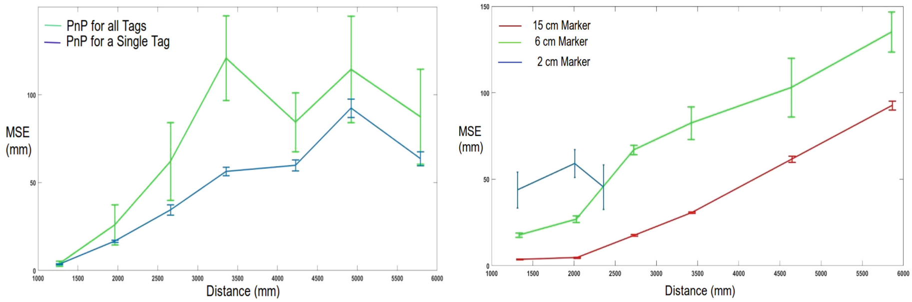
Fig. 9. Experimental set-up used for the localisation of a nano-drone CrazyFly.

Figure 10 presents an evaluation of the estimation error for the position obtained by solving the PnP problem. Figure 10a compares two estimation approaches when multiple tags are detected for the same object of interest. The first solution (green) solves the PnP problem for each tag separately and then calculates the object location as the average of the detected tags positions. In the second approach, all the detected tags are combined in a single vector and the PnP problem is then solved for the combined tags.

Figure 10b depicts the performance of the position estimation by solving the PnP problem using different sizes of the marking tags. As was expected, the precision of the estimation increases with the size of the tag. The detection algorithm was incapable of detecting the small tag, from distances exceeding two meters with an error around 5 cm for such a distance. The biggest tag on the other hand, was detectable even at a distance of more than 6 m, with an average error of less than 1 cm up to a depth of two meters, and less than 9 cm for a distance not exceeding 10 m.

Figure 11a presents a comparison of the precision of the triangulation and PnP methods. As this figure shows, the triangulation algorithm gave better estimation for all the tested positions in the scene.

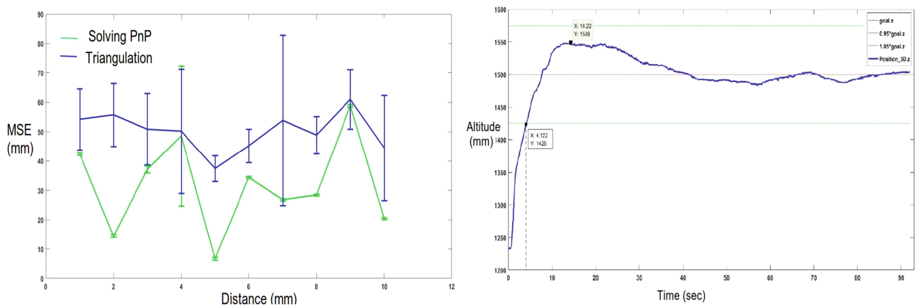
To test the solution on a real platform the nano-drone CrazyFly was used. The latter, although having a very limited payload of merely 4 g, was able to



(a) (b)

Fig. 10. Position estimation error using the PnP solution: (a) Solving PnP for all detected tags vs. solving PnP for each tag separately; (b) Estimation Precision for different tag sizes. (Color figure online)

carry a cube of 4 tags as well as the coloured marker. The obtained position was used as a control feedback to stabilise the altitude of the vehicle. In order to achieve that, the localisation solution was integrated to the robotic operating system (ROS), which facilitated the interface to the drone. Figure 11b depicts the obtained preliminary results using a discrete PID controller implemented on the same laptop running the localisation algorithms.



(a) (b)

Fig. 11. Experimentation results: (a) Comparison of position estimation precision using triangulation and PnP methods; (b) Results of altitude visual servoing for the nano-drone CrazyFly.

5 Conclusion

Motivated by the actual high cost of motion tracking systems, this work targeted the development of a low-cost localisation solution for indoor environments. The proposed solution was based on a multi-camera system. We showed, through the experiments conducted, that a system composed of four ordinary IP cameras, can provide a precision of a couple of centimetres up to a distance of two meters. Challenges related to the real-time operation of the localisation system were overcome using the adequate implementation. This allowed processing the images of the four cameras in full rate on an ordinary laptop.

Future works target a more thorough evaluation of the positioning errors, in order to achieve an optimal fusion of the measurement obtained from the four cameras. Another aspect, which should enhance the precision and further reduce the processing time, is the introduction of a filtering approach which takes into account the previous position of the object of interest in both the detection and position estimation steps.

References

1. Caron, F., Duflos, E., Pomorski, D., Vanheeghe, P.: GPS/IMU data fusion using multisensor Kalman filtering: introduction of contextual aspects. Inf. Fusion **7**(2), 1–19 (2006). http://www.sciencedirect.com/science/article/pii/S156625350400065X
2. Fiala, M.: ARTag, a fiducial marker system using digital techniques. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 590–596 (2005). http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1467495
3. Groves, P.D.: Principles of GNSS, Inertial, and Multisensor Integrated. Artech House, California (2008)
4. Kato, H., Billinghurst, M.: Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In: Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR 1999), pp. 85–94 (1999)
5. Kendoul, F.: Survey of advances in guidance, navigation, and control of unmanned rotorcraft systems. J. Field Robot. **29**(2), 315–378 (2012). https://doi.org/10.1002/rob. http://onlinelibrary.wiley.com/doi/10.1002/rob.20414/full
6. Lee, G.H., Achtelik, M., Fraundorfer, F., Pollefeys, M., Siegwart, R.: A benchmarking tool for MAV visual pose estimation. In: 11th International Conference on Control, Automation, Robotics and Vision, ICARCV 2010, vol. 1, pp. 1541–1546 (2010). https://doi.org/10.1109/ICARCV.2010.5707339
7. Lochmatter, T., Roduit, P., Cianci, C., Correll, N., Jacot, J., Martinoli, A.: SwisTrack-a flexible open source tracking software for multi-agent systems. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2008, pp. 4004–4010 (2008)
8. Oh, H., Won, D.Y., Huh, S.S., Shim, D.H., Tahk, M.J., Tsourdos, A.: Indoor UAV control using multi-camera visual feedback. J. Intell. Robot. Syst. Theory Appl. **61**(1–4), 57–84 (2011). https://doi.org/10.1007/s10846-010-9506-8
9. Olson, E.: AprilTag: a robust and flexible visual fiducial system. In: IEEE International Conference on Robotics and Automation, pp. 3400–3407. IEEE, May 2011. https://doi.org/10.1109/ICRA.2011.5979561. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5979561
10. Visvanathan, R., et al.: Mobile robot localization system using multiple ceiling mounted cameras. In: IEEE SENSORS, pp. 1–4 (2015)
11. Schmidt, A., Kasiński, A., Kraft, M., Fularz, M., Domagała, Z.: Calibration of the multi-camera registration system for visual navigation benchmarking. Int. J. Adv. Rob. Syst. **11**(1) (2014). https://doi.org/10.5772/58471

12. Shim, J.H., Cho, Y.I.: A mobile robot localization using external surveillance cameras at indoor. Procedia Comput. Sci. **56**, 502–507 (2015). https://doi.org/10.1016/j.procs.2015.07.242. http://linkinghub.elsevier.com/retrieve/pii/S1877050915017238
13. Wagner, D., Schmalstieg, D.: Making augmented reality practical on mobile phones, Part 2. IEEE Comput. Graphics Appl. **29**, 6–9 (2009). http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5167481