METHODOLOGY

Precision mapping through the stereo vision and geometric transformations in unknown environments

Georgios Petrakis[1] · Angelos Antonopoulos[1] · Achilles Tripolitsiotis[2] · Dimitris Trigkakis[1] · Panagiotis Partsinevelos[1]

Abstract

Geodetic techniques for surveying and rapid mapping need to be revisited due to the present progress on satellite, sensor, and geospatial technologies. Conventional surveying methods provide high level of accuracy but require significant human involvement in the field while GNSS (Global Navigation Satellite System) positioning method, provides unsatisfactory accuracy in urban or high vegetated areas due to the degraded GNSS signal coverage. In this study, an alternative surveying method is proposed which facilitates the process of characteristic point localization, using stereoscopic vision and at least one visual marker. At first, the camera system localizes itself and maps the study area using stereo SLAM (Simultaneous Localization and Mapping) algorithm while subsequently detects the visual markers (origin and targets), placed in the area. Afterwards, using a multi-view geometry method for the marker localization and an optimization algorithm for origin marker's plane alignment, the system is able to export the coordinates of the markers and a point cloud (provided by SLAM) in a local coordinate system based on the origin marker's pose and location. The study involves both terrestrial and unmanned aerial vehicle platforms that may carry the proposed equipment. An extensive set of indoor and outdoor, terrestrial and UAV experiments validates the methodology which succeeds a horizontal and vertical error in a level of 10 cm or better. To the best of our knowledge this study proposes the first surveying alternative which requires only a stereo camera and at least one visual marker in order to localize specific and arbitrary points in a centimeter level of accuracy. The proposed methodology, demonstrates that the use of low-cost equipment instead of the costly and complicated surveying equipment, may prove sufficient to produce an accurate 3D map of the scene in an unknown environment.

Keywords Stereo vision · Localization · SLAM · Unmanned aerial vehicle · Surveying

Introduction

Geospatial technology, based on higher geodesy, remote sensing and geographical information science has change the way that scientists, engineers and citizens study or interact with their environment. This fact results in fundamental advances in various topics of geomatics such as location-based applications, spatial data infrastructures, navigation or geodetic equipment. Nevertheless, conventional surveying, although is the most accurate and robust method of applied geodeysy, it remains a time consuming process with significant human effort (Carrera-Hernández et al. 2020). On the other hand, GNSS (Global Navigation Satellite System) positioning method, provides unsatisfactory accuracy in urban or high vegetated areas due to the degraded GNSS signal coverage (Chiang et al. 2019).

Specifically, traditional and modern surveying methods are not always complementary, since in many cases, the use of total stations is mandatory and cannot be substituted with GNSS receivers, while in other cases traditional surveying is prohibitive. However, most surveying procedures including laser scanning involve costly equipment while the necessity of a cost-effective surveying alternative in GNSS-degraded environments remains a critical unresolved issue (Trigkakis et al. 2020).

In the initial research of this study (Trigkakis et al. 2020), alternative solutions for positioning in GNSS-degraded areas are presented. Some approaches involve the improvement

✉ Georgios Petrakis
gpetrakis2@tuc.gr

Achilles Tripolitsiotis
admin@spacegeomatica.com

1 SenseLab Research, Technical University of Crete, GR-73100 Chania, Greece

2 Space Geomatica P.C., Xanthoudidou 10A, GR-73132 Chania, Greece

of signal with respect to independent system analysis (Panigrahi et al. 2015) while other methodologies propose alternative techniques including angle approximation (Tang et al. 2015), shadow matching (Urzua et al. 2017), multipath estimations using 3D models (Zahran et al 2018) and statistical models (Romero-Ramirez et al. 2018; Partsinevelos et al. 2020). Various studies make use of high-resolution aerial or terrestrial images and statistical / machine learning algorithms in order to georeference, map and detect dynamic patterns (Zahran et al. 2018; Jende et al. 2018), while the use of simultaneous localization and mapping (SLAM) algorithms combined with complementary methods from Photogrammetry and / or GNSS / INS (Inertal Navigation System) for localization in GNSS denied environments is signified in the literature (Bobbe et al. 2017; Gabrlik 2015; Helgesen et al. 2019). Under the same perspective, some studies use monocular SLAM approaches and attempt to solve the inherent problem of scale estimation by using barometers, altimeters and landmarks (Urzua et al. 2017; Kuroswiski et al. 2018) or by utilizing orientation (AHRS) and position sensors (GPS) (Munguía et al. 2016). In Lichao Xu et al. (2019), a localization method for indoor environments is presented, which is able to recognize pre-defined markers with centimeter level of accuracy utilizing RGB-Depth ORB-SLAM2 algorithm.

Several studies have been conducted for accurate and / or rapid mapping with the use of mobile mapping systems (MMS), Photogrammetry and image processing techniques. In Kalacska et al. (2020), authors follow the approach of Structure-from-Motion (SfM) with multi-view stereo technique of Photogrammetry to produce ortho-images and 3D surfaces without the use of ground control points (GCPs) using UAVs equipped with GNSS receivers and optical sensors. In Pinto and Matos (2020), densely 3D information in underwater environments is constructed through the fusion of multiple light stripe range (LSR) and photometric stereo (PS) methods outperforming the corresponding conventional methods in terms of accuracy while in Bañón et al. (2019), aerial images and ground control points (GCPs) are used in order to produce a 3D model in a coastal region through SfM. The characteristic points are measured using a GPS receiver for the validation of the methodology with a vertical RMSE error of 0.12 m. Tomaštík et al. (2017), evaluate the positional accuracy of forest rapid—mapping, using point cloud data created by UAV images and the Agisoft software with an accuracy level of 20 cm.

Various studies are referred to localization and detection methods employing MMS equipped with stereo sensors. Haque et al. (2020), propose an unmanned aerial system (UAS) which is able to find its location in a 3D CAD model of a pre-defined environment. The UAS with a stereo-depth camera, maps the area using OrbSLAM2 algorithm (Mur-Artal and Tardós 2017), detects and extracts vector features with the aid of a convolutional neural network (CNN) and rectifies its location comparing the SLAM mapping area with the 3D CAD model. In Li et al. (2017), authors propose a pose estimation methodology based on mobile accelerometers, visual markers and stereo vision fusion, achieving a centimeter level of accuracy while in (Vrba and Saska 2020; Vrba et al. 2019), a methodology that detects a micro aerial vehicle (MAV) is proposed, utilizing machine learning techniques and an RGB-Stereo depth camera with an average RMS error of 2.86 m. In Zhang et al. (2019), a real-time obstacle avoidance method is developed with the aid of a stereo camera, a GNSS receiver and an embedded system mounted on a UAV in order to detect obstacles and follow an alternative, obstacle-free path. In Ma et al. (2021), authors utilize a UAV with two cameras and a GNSS receiver in order to detect and geographically localize insulators in power transmission lines based on the bounding box of the detected insulators. Moreover, stereo-depth cameras have been used in UAVs for autonomous landing in GNSS-denied environments, where a UAV is able to detect, locate and land on an unmanned ground vehicle (UGV) making use of information from a multi-camera system and deep learning algorithms (Yang et al. 2018; Animesh et al. 2019).

As referred above, the literature abounds of positioning methodologies for GNSS-denied areas, rapid mapping solutions using photogrammetric techniques or localization systems based on SLAM and detection. Although most of the studies propose alternative localization solutions, none of them focus on surveying or traditional topography combined with computer vision and multi-view geometry algorithms.

In the monocular approach of the present methodology (Trigkakis et al. 2020) an implementation based on SLAM, point cloud and image processing techniques, localizes characteristic points in a local coordinate system utilizing only a monocular camera attached on a UAV in combination with a visual marker. Although the main issue of the monocular setup approaches is the scale estimation (Sahoo et al. 2021), the proposed methodology controlled this issue by using the "Multiple convergence" method achieving an accuracy level of 50 cm (Trigkakis et al. 2020).

In this study, the methodology is further extended using a stereo camera instead of a single sensor and validated conducting various indoor and outdoor, UAV and terrestrial experiments. More specifically, the extended methodology takes advantage of a stereo camera and a visual marker, and is capable to map an unknown area, providing refined estimations of point coordinates in a local 3D coordinate system fusing stereo SLAM, image processing techniques and coordinate system transformations. The main objective of this study, is to propose a surveying or rapid-mapping alternative with an accuracy level of 10 cm or better, using conventional components, while supporting the use of a UAV. Although the present study validates the methodology in terms of localization accuracy in a local coordinate system,

the use and connection with a global coordinate reference system such as WGS-84 is quite possible. The methodology can be employed in urban environments or dense-canopy areas where the GNSS signal is degraded, in emergency situations (Chuang 2018) where there is a need for damage assessment (Ampadu et al. 2020; Lassila 2018) and / or in search and rescue applications (Mishra et al. 2020; McRae et al. 2019). The proposed methodology is a cost-effective, rapid and efficient surveying solution where a few minutes of flying and processing are sufficient to map an area of interest and extract the coordinates of the characteristic points without limitations related with steeply slope terrains or occluded areas.

To the best of our knowledge, there is no similar solution that makes use of a visual SLAM algorithm, a stereo camera and a visual marker in order to provide a 3D local coordinate system in a 10 cm level of accuracy. Unlike the similar localization methods, the coordinate estimations were not extracted in a software-based reference system but in a reference system which is well-defined in the scene. The main contribution of the study are as follows:

- An alternative surveying solution was developed using stereo-SLAM, multi-view geometry and coordinate system transformations.
- The methodology can be performed with minimum and cost-effective equipment since a stereo-camera and at least one visual marker are enough to map an unknown environment localizing characteristic points in a 3D local coordinate system.
- All coordinate estimations are transferred and exported in a local reference system which is well-defined in the scene, using the plane and the pose of a visual marker.
- The proposed solution provides an accuracy in a level of 10 cm or better, a significant improvement compared with the monocular approach.

In the following sections the core system, the equipment and the implementation are presented, while in Section 3 an extensive set of experiments and results validate the presented methodology and demonstrate that can be used as an alternative surveying solution. Sections 4 and 5 discuss the results and conclusions of the proposed methodology.

Materials and methods

The main goal of this study is the localization of visual markers and characteristic points of the scene, providing their local coordinates in 3D space under a high level of accuracy using minimal equipment. The presented methodology maps the area of interest, by extracting the pose estimation of pre-defined visual markers and a point cloud

in a local coordinate system using stereo vision. At first, the visual markers are placed in the scene; the origin marker defines the reference system of the coordinate estimations while the target markers are represent the characteristic points or features. Subsequently, a SLAM algorithm enables the stereo camera to map the desired area and localize itself in an unknown environment (Mur-Artal and Tardós 2017), while in combination with image and geometrical processing, the present methodology estimates the coordinates of target markers and an arbitrary point cloud which approximates the structure of the environment, allowing additional measurements in the local coordinate system of the scene.
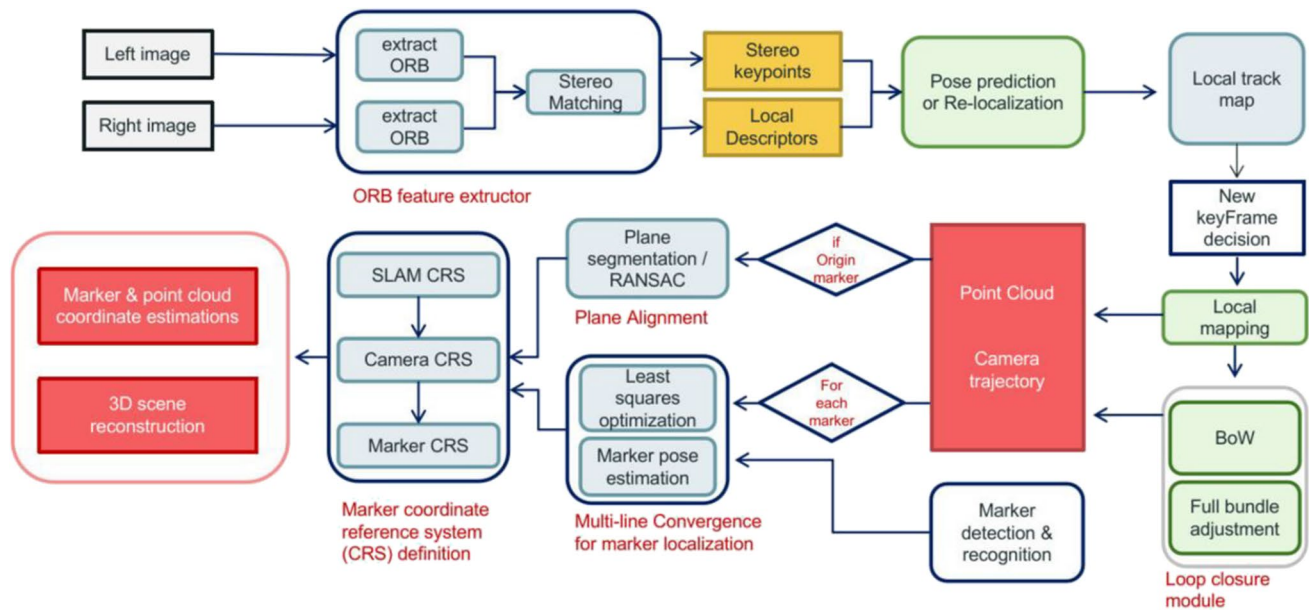
System architecture

The system architecture is presented in the following figure (Fig. 1).

As presented in Fig. 11, the processing levels in the procedure are separated in three stages. Initially, a video is captured by a stereo camera creating a bag file (http://wiki.ros.org/Bags). After the recording process, scripts in Python language extract stereo imaging data using ROS (Robot Operating System, https://www.ros.org/), obtain calibration information utilizing camera's factory settings while separate and store the imaging data per sensor.

Subsequently, the SLAM uses ORB (Rublee et al. 2011) algorithm in order to extract the interest keypoints of images and the local descriptors which aid the system to recognize the features from multiple angles and distances. The SLAM algorithm extracts the ORB features from both images (left and right) while for each ORB feature of the left image, detects the corresponding ORB feature of the right image. The coordinates of stereo interest keypoints are defined as (uL,υL, uR) (Mur-Artal and Tardós 2017) where the first two coordinates (uL,υL) are the horizontal and vertical coordinate of the the left image while the third one is the horizontal coordinate of the right image. Afterwards, the system, using the internal parameters of the camera and the information of the extracted features, predicts the position and orientation of the camera (pose), while if it observes groups of features in multiple sequential frames, it stores a keyframe.

Based on the process above, the SLAM algorithm outputs multiple keyframes which are treated as landmarks since, in combination with the keypoints, are necessary for the local mapping, the loop closure detection and for the re-localization of the camera. For the optimization of the camera's pose prediction, local mapping and loop closure detection, SLAM algorithm utilizes the bundle adjustment (BA) algorithm using the Levenberg–Marquardt method (Mur-Artal and Tardós 2017).

After the end of the SLAM process, it outputs a point cloud and a trajectory of the scene while traditional image processing techniques such as adaptive and Otsu

Fig. 1 System architecture of the proposed methodology

thresholding (Otsu 1979) provide the identifications of target markers. Subsequently, through the multi-line convergence method (Trigkakis et al. 2020), the locations of the markers are estimated while the pose of the origin marker is optimized with the utilization of plane alignment method (Trigkakis et al. 2020). Finally, the coordinate estimations are transferred in a local coordinate system, defined by the pose of the origin marker (see Section 2.2). After the end of the process, the resulted 3D scene with the point cloud, the camera trajectory and the marker estimations, are graphically presented through the visualization module (see Section 2.3).

The study's methodology performs SLAM processing based on ORB-SLAM2 (Mur-Artal and Tardós 2017) algorithm making use of two infrared sensors, while ArUco library (Romero-Ramirez et al. 2018; Garrido-Jurado et al. 2016) is utilized to represent origin and target markers. ArUco markers, are synthetic square markers defined by a binary matrix (black and white) with a black border and a specific identifier (id), meaning that different markers are given different identities.

The ORB-SLAM2 algorithm was selected due to its robustness over several state-of-the-art SLAM alternative solutions such as LDSO (Gao et al. 2018), openVINS (Geneva et al. 2020), VINS-fusion (Qin et al. 2019), Maplab (Schneider et al. 2018), Basalt (Usenko et al. 2019), Kimera (Rosinol et al. 2020) and open-VSLAM (Sumikura et al. 2019). In Sharafutdinov et al. (2021), authors compare the above SLAM alternatives using the ablsolute pose error metric in position and rotation and quite popular datasets in robotics such as EuRoC MAV (Burri et al. 2016),

OpenLoris-Scene (Shi et al. 2020) and KITTI (Geiger et al. 2012). The results prove that ORB-SLAM2 and openVS-LAM achieved the highest overall accuracy. Moreover, in (Giubilato et al. 2018), authors compare stereo visual SLAM algorithms for planetary rovers proving the superiority of ORB-SLAM2 against the S-PTAM (Pire et al. 2015), LibVISO2 (Geiger et al. 2011), RTAB-MAP (Labb and Michaud 2013) and ZED-VO (the proprietary software by ZED development toolkit, https://www.stereolabs.com/developers/release/).

ORB-SLAM2 in stereo mode provides a real-world scale that is given in meters due to the known camera baseline between the two sensors instead of the monocular solution in the previous version of this study (Trigkakis et al. 2020) where the scale was calculated mathematically.

Functionality

A core component of the methodology for the final coordinate estimations and 3D scene reconstruction is the coordinate system definition. The first coordinate system is defined and established by ORB-SLAM2 using the first frame of the captured video. The x and y axes in this initial coordinate system, follow the right and top directions of the frame respectively while the z axis is equivalent with the camera direction towards the landscape of the area. Subsequently, the calibration data and the camera pose (retrieved by camera trajectory information) along with the target marker coordinates which are calculated by OpenCV (https://opencv.org/) algorithms, extract the vectors of rotation and translation that are utilized in transformation of the initial reference system to the camera

reference system. Finally, the reference system definition module, calculates the translation vector and rotation matrix from the orientation and translation of the origin marker and defines the final reference system based on the origin marker's pose. The x and y axes of the marker reference system follow the right and top direction of the marker while the z axis follows the zenith direction as depicted in Fig. 2 below.

Concerning the marker pose estimates, the Multiple Line Convergence (M.L.C.) method (Trigkakis et al. 2020) was implemented. M.L.C. is a method for the marker location definition that is based on the observation that the extended line segments which connect each pose estimate with corresponding camera position, converge in an area that corresponds to the location of the marker in the 3D scene. The method defines the optimized point that the extended line segments converge, using pseudo-inverse least squares optimization (Samuel 2004; Eldén 1982). The method can be described by the following equation:
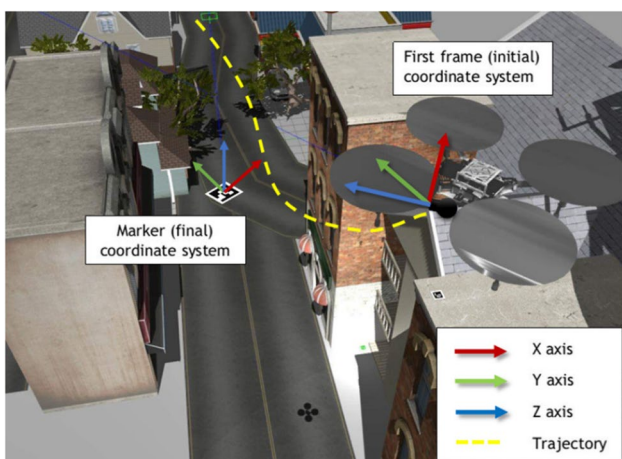
$$p = S^{+} \cdot C \tag{1}$$

where p is the minimized distance of the theoretical convergence point from all the lines while S + is the pseudo-inverse matrix of S which is defined in Eq. 2. C is defined in Eq. 3.

$$S = \sum_{i} \left[n_i n_i^T - I \right] \tag{2}$$

$$C = \sum_{i} \left[n_i n_i^T - I \right] a_i \tag{3}$$

where each line is defined with "i", "α_i" is the starting point of line "i" and "n_i" is the direction of line i while "I" is an identity matrix.

Fig. 2 Coordinate system transformation. During the mappping process, the proposed methodology uses the first recorded frame as defined by ORB-SLAM2 while after the export of coordinate estimations, the system defines a local coordinate system using the pose of the origin marker. Image source: (Trigkakis et al. 2020)

Subsequently, Plane alignment method (P.A.) (Trigkakis et al. 2020) is performed to correct the translation and rotation errors of the origin marker that defines the final coordinate system of the scene. This step is important because any pose estimation error in the origin marker is transferred in every target marker and point cloud data of the scene. With the P.A. method, the pose and rotation of the origin marker is corrected leading to reliable measurements and an accurate definition of the origin coordinate system.

Visualization module

The visualization of the results is a requirement during the implementation, testing and experimentation phases of the methodology since visualizing the data in a 3D scene is important for tasks that require metric information from the corresponding real-world scene. Beyond the visualization capabilities that the module provides through the processing, it can be used in offline mode with a built-in user interface, allowing the navigation and interaction within the visual scene. More specifically, it is able to render the entire scene containing the marker estimations, the camera trajectory and the point cloud, while the vectors of the camera trajectory pose and the detection line segments of the camera to the markers' center are also depicted in the 3D scene (Fig. 3).

From a more practical perspective, the user interface supports point selection through a cursor extracting the corresponding coordinate estimations on the screen, while providing a succinct overview of the camera trajectory reproducing the path of the camera through animation.

Figure 3 depicts the user interface (UI) of the visualization module in which a part of a 3d scene is visualized. All the features that are displayed in the figure are located in the local reference system based on the origin marker. The poses (translation and orientation) of the camera are depicted using the line segments (in green, Fig. 3) which follow the trajectory of the camera while the marker detections of the camera are visualized as red lines which connect each camera position with center of the detected marker. It's worth mentioning that the center of markers is placed on the convergence point of the lines based on the multi-line convergence method. Finally, the user is able to select a point from the point cloud that are visualized in green color, exporting on the top-right of the screen the coordinates from the initial and final coordinate system while if more points are selected, the module exports the corresponding mean values of the selected point.

Equipment setup

The main equipment components include a stereo camera, the aruco markers, a conventional computing system and a UAV. Concerning the stereo-camera, the Intel Realsense D435

Fig. 3 Multiple point selection inside the visualization module. A point is then shown through the UI at the top of the screen, in its original point cloud coordinates, as well as with coordinates expressed in a marker-determined coordinate system. The same process applies to the selected points, by averaging the location and treating it as a new point, with both original and marker-determined coordinates
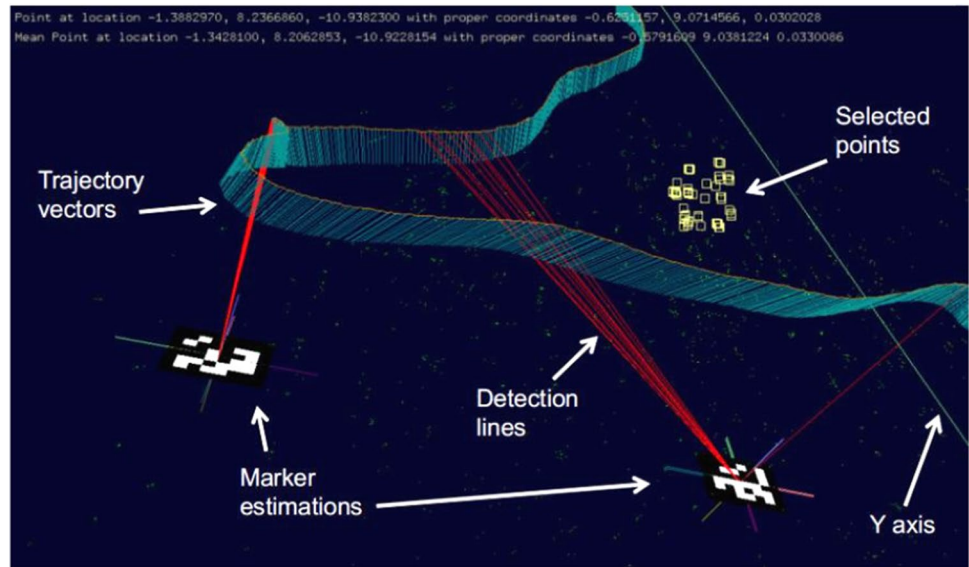
stereo-depth camera was used which includes two infrared sensors (left and right), a color sensor and an infrared projector for the depth information. In the present study, only the two infrared sensors were used. The resolution of the camera sensors is 1280×800, the sensor aspect ratio is 8:5, the focal length 1.93 mm, while the format is 10-bit RAW.

The UAV that was used in the present study is a custom-made hexacopter (Fig. 4a). The frame as well as the propellers, are made of carbon fiber while the Flight Control Unit (FCU) is a Pixhawk 2 Cube. The UAV is designed to be used with a companion embedded computer, a Raspberry Pi 4 module with 8 GB of RAM running at an overclocked rate of 2.3 GHz (Fig. 4b). The embedded computer interfaces with both the FCU as well as the Realsense camera (Fig. 4c).

The origin and the target markers have a size of 30×30 cm while they are installed in a custom-made adjustable stand. This stand is able to stabilize the marker pose in a horizontal reference plane with the aid of two stainless steel threaded rods and a leveler (Fig. 5a).

For validation purposes and ground-truth measurements, a Topcon GPT 3000 geodetic total station was used with $\pm (3 \text{ mm} + 2 \text{ ppm} \times D)$ mean square error (m.s.e) measurement accuracy where D is the measured distance between the total station and the prism (Fig. 5b).

Results

To validate the present methodology an extended set of experiments was performed under different conditions relating to the study area, the arrangement of markers on the ground and the use or not of a UAV.

For the evaluation process, a geodetic total station was utilized in order to measure the reference coordinates of the visual markers and several characteristic points. The origin of the local coordinate system was defined using the center of the origin marker with the coordinates $X = 0$, $Y = 0$ and $Z = 0$. It's worth mentioning that the videos were recorded at 90fps using 848×480 resolution.

(a) (b) (c)
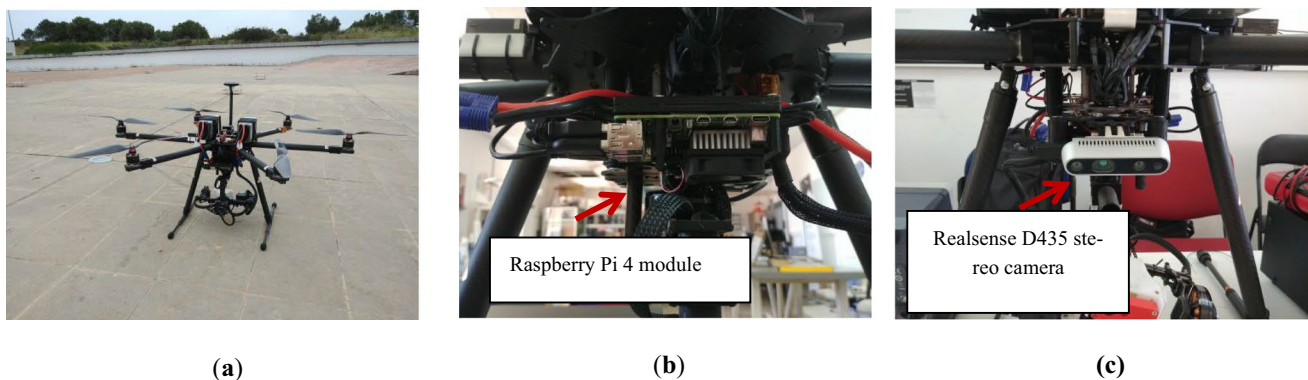
Fig. 4 **a** The custom-made hexacopter UAV. **b** The Intel Realsense D435 stereo camera attached on the UAV. **c** The Raspberry Pi 4 module

Fig. 5 **a** The origin marker located on a custom-made adjustable stand which is able to stabilize the marker pose in a horizontal reference plane using two stainless steel threaded rods and a leveler. **b** The GTP-3000 geodetic total station
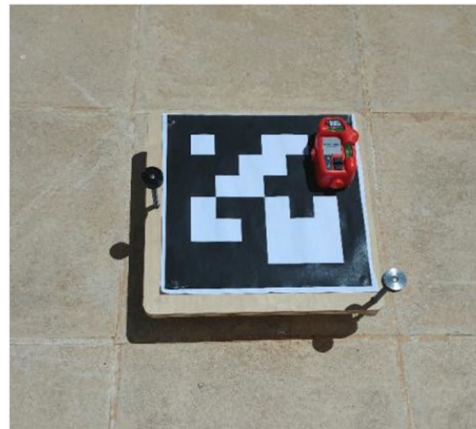
(a)　　　　　　　　　　**(b)**

For the evaluation of the experiments the absolute error ($|X_{meas} - X_{est}|$) between the measured coordinates of X, Y, Z and the corresponding estimations is used while the horizontal error ($\sqrt{X_{err}^2 + Y_{err}^2}$) is also calculated. The experiments below are separated in two main sections: terrestrial and UAV experiments.

In each experiment, a visual marker which represents the origin of local coordinate system and one or three markers which represent the targets are located to the scene and measured with the total station for ground truth information. Afterwards, the stereo camera, stand-alone or attached on the UAV is guided through a desired trajectory path in order to identify the markers and maps the surroundings.

The experiments were designed aiming to simulate a real-case scenario of surveying a plot or a field in which traditional land surveying techniques and equipment are utilized. More specifically, the main field-work of a surveyor is to measure the coordinates of a few points that form the borders of the mapping area while in most of the cases, the path that the surveyor follows can be approached with straight, right-angle, step-shaped, pi-shpaed and squared-based paths.

Thus, in the present experimentation, the methodology was tested utilizing the commonly-used paths that referrred above while the visual markers which represent the characteristic points of the path were placed on locations aiming to form the shape of each path similarly to a real-case scenario. For instance in a surveyed area with square shape, the visual markers are placed in the four corners of the square while in a long straight path, two or more visual markers are located along the straignt path.

Terrestrial experiments

Straight path (indoor)

The first experiment was conducted in an indoor environment. Two markers were utilized for the origin and target marker respectively, in a distance of 3.5 m while the camera followed a straight path as presented in Fig. 6a. It's worth mentioning that the target marker was placed about 70 cm higher than the origin marker. The results of this experiment are presented in Table 1.

As it's presented in Table 1, the X error is below 5.50 cm while Y error below 3 cm with the horizontal error in a level of 6 cm while the vertical error is below 4 cm.

Right-angle path (indoor)

This experiment was performed in an indoor environment. Three target markers and an origin marker were used, placed in a right-angle shape (Fig. 6b). The maximum distance between the origin and the farthest target marker is 7.20 m.

In this experiment, as it is observed in Table 2, a horizontal error of 2.57 cm in target 1 with 0.019 cm of error in X axis and 2.569 cm in Y axis while in targets 2 and 3 the horizontal error reaches a level of 6 cm with about 5.50 cm of error in X axis and 5 cm in Y axis. The vertical error in target 1 is 3.10 cm while in the targets 2 and 3 the accuracy is decreased with an error of 9 cm.

Step-shaped path (indoor)

In the following indoor experiment, three target markers and an origin marker were used placed in the same shape as in previous experiment but the camera followed a step-shaped path (Fig. 6c). The maximum distance between the origin and the final target marker is 7.20 m.

In Table 3, the horizontal error of target 1 is about 2 cm with 0.85 cm of error in X axis and 1.9 cm in Y axis while in targets 2 and 3, an accuracy below 5.5 cm is observed. In target 2 the X and Y errors are about of 3 cm while target 3 reaches an error of 5 cm in X and 2.3 cm in Y. The

Fig. 6 Trajectory paths of indoor experiments **a** straight path **b** right angle path **c** step-shaped path
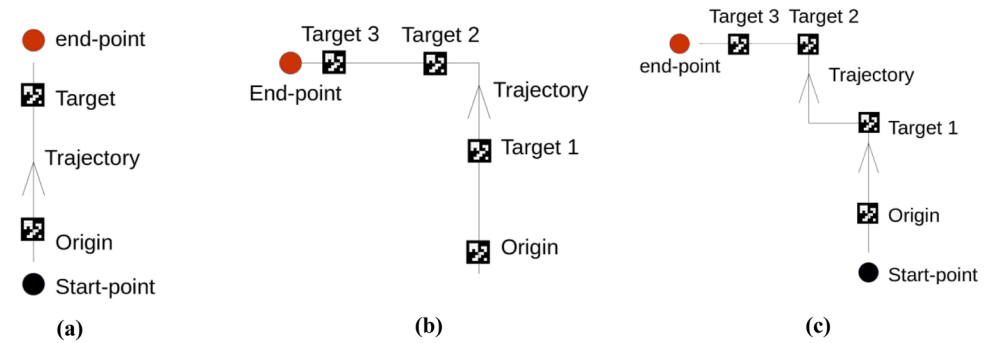
Table 1 Estimations of straight path (indoor) experiment

Target marker	X (cm)	Y (cm)	Z (cm)
Ground Truth	-48.0	350.0	70.60
Estimations	-42.71	347.29	66.80
Error	5.29	2.71	3.70
XY error	5.94		-

Table 3 Estimations of step-shaped path (indoor) experiment

Target marker	X (cm)	Y (cm)	Z (cm)
Target 1: Ground Truth	0	340.5	0
Target 1: Estimations	0.85	338.6	0.08
Target 1: Error	0.85	1.9	0.08
Target 1: XY error	2.08		-
Target 2: Ground truth	-89.0	690.5	0
Target 2: Estimation	-85.244	687.3	5.14
Target 2: Error	3.156	3.20	5.14
Target 2: XY Error	4.49		-
Target 3: Ground truth	-212.50	690.50	12.0
Target 3: Estimation	-207.662	688.275	19.31
Target 3: Error	4.838	2.225	7.31
Target 3: XY Error	5.33		-

Table 2 Estimations of right-angle path (indoor) experiment

Target marker	X (cm)	Y (cm)	Z (cm)
Target 1: Ground Truth	0	340.5	0
Target 1: Estimations	-0.019	337.931	3.10
Target 1: Error	0.019	2.569	3.10
Target 1: XY error	2.57		-
Target 2: Ground truth	-89.0	690.5	0
Target 2: Estimation	-86.711	685.719	9.062
Target 2: Error	2.289	4.781	9.06
Target 2: XY Error	5.30		
Target 3: Ground truth	-212.50	690.50	12.0
Target 3: Estimation	-209.212	685.49	20.783
Target 3: Error	3.288	5.01	8.783
Target 3: XY Error	5.99		

was wider than the areas of indoor environments and the distance between the origin and the farthest marker was 10 m larger than the experiments 2 and 3, the horizontal accuracy was higher. This result is reasonable because in a sunny outdoor environment the sensors receive refined information because of illumination conditions of the scene, which results in more accurate mapping while the geometry of the outdoor scene such as buildings and trees aids the methodology to provide more accurate results for target markers. Concerning the vertical error, the targets 1 and 2 succeed an error below 7.8 and 2.5 cm respectively while in target 3, the error is increased in a level of 25 cm. This difference of the vertical error between target 3 and targets 1—2 is due to the variable height of the camera between target 2 and target 3 because of different altitude levels of the ground. This is an interesting result which determines that the camera height should be stable during the mapping process which is proved with the UAV-experiment results in the following section.

UAV experiments

Straight path

The first experiment with the custom-built UAV was conducted in an outdoor environment without obstacles such as

vertical error of target 1 is below 1 mm while in targets 2 and 3 the error is about 5.50 cm and 7.50 cm respectively.

Pi-shaped path (indoor)

This experiment was conducted in an outdoor environment in a sunny day. Three markers were used as targets which were placed in a pi-shaped path as depicted in Fig. 7. The maximum distance between the origin and the target 3 marker was about 17.50 m.

In this outdoor experiment, as it is presented in Table 4, the horizontal error of target 1 is 1.70 cm with 0.08 cm of error in Y axis while target's 2 horizontal error is below 3.5 cm with 0.4 cm of error in X axis. In target 3 the error is increased in a level of 9 cm with 0.3 cm of error in X axis and 8.8 cm in Y axis. As observed, although the study area in this experiment

Fig. 7 **a** Survey area **b** Right angle trajectory path

(a)   (b)

Table 4 Estimations of pi-shaped path (outdoor) experiment

Target marker	X (cm)	Y (cm)	Z (cm)
Target 1: Ground truth	244.4	0	-6.0
Target 1: Estimation	242.7	0.08	1.8
Target 1: Error	1.7	0.08	7.8
Target 1: XY Error	1.70		-
Target 2: Ground truth	105.6	920.2	-4.0
Target 2: Estimation	105.2	916.8	-1.5
Target 2: Error	0.4	3.4	2.5
Target 2: XY Error	3.42		-
Target 3: Ground truth	621.7	1618.1	-24.0
Target 3: Estimation	621.4	1609.3	0.8
Target 3: Error	0.3	8.8	24.80
Target 3: XY Error	8.81		-

is refined compared with the non-UAV indoor and outdoor experiment due to stable height of the camera.

Square path

The last experiment was performed in the same area of the previous experiment (Fig. 8a) using the custom-made UAV. Four markers were used, one for the origin and three for the targets in a distance of about 5 m while the UAV followed a square path with a flight height of 3 m (Fig. 9). This experiment was repeated 50 times under similar conditions and equipment setup in a period of a month in order to evaluate the presented methodology using statistical metrics. For each target marker the absolute error between the coordinate estimations (X, Y, Z) and the corresponding ground truth

trees or buildings. Two markers were utilized for the origin and target marker respectively in a distance of about 5 m while the UAV followed a straight path with a flight height of 3 m. (Fig. 8).

As presented in Table 5 the horizontal error is 7.15 with below 5.50 cm of error in X and Y axis while the vertical error is 1.15 cm. It's worth to mention that the vertical error

Table 5 Estimations of straight path (outdoor) experiment

Target marker	X (cm)	Y (cm)	Z (cm)
Ground Truth	0.00	486.00	0.2
Estimations	-4.74	480.65	-1.15
Error	4.74	5.35	1.35
XY error	7.15		-

Fig. 8 **a** Survey area **b** Straight trajectory path

(a)   (b)
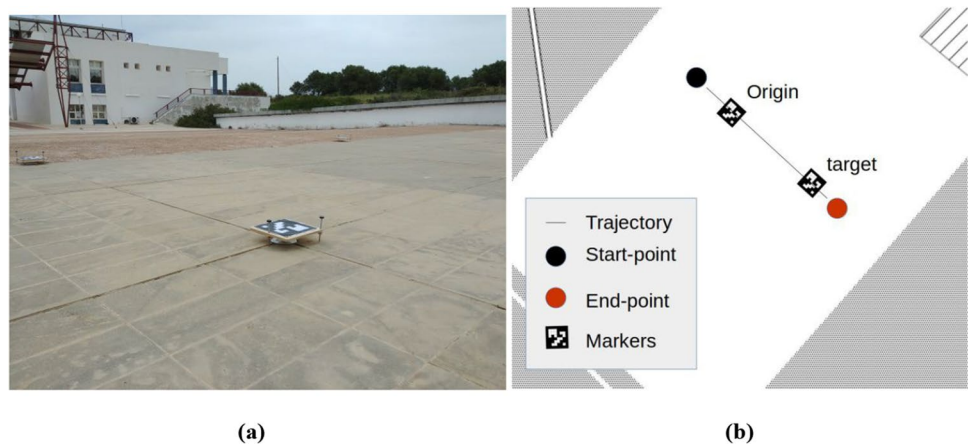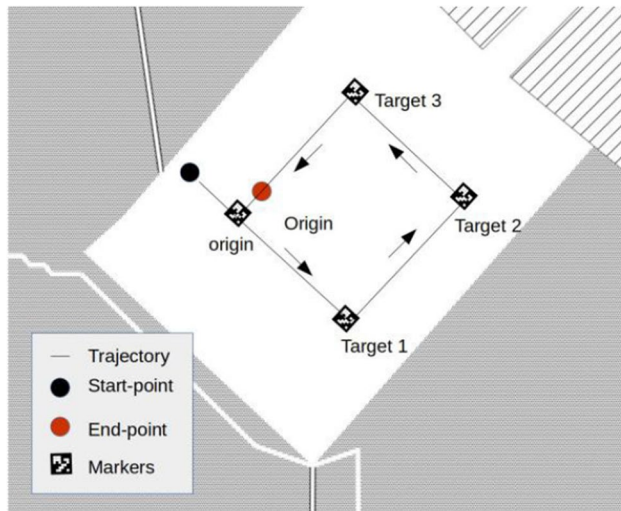
Fig. 9 Square trajectory path and camera direction

Table 6 Evaluation of square path (outdoor) experiment for target 1

Target 1	X (cm)	Y (cm)	Z (cm)
Average	4.50	7.96	1.80
Std. Deviation	2.93	4.89	1.23
RMSE	5.27	9.19	2.13
RMSE xy	10.59		

Table 7 Evaluation of square path (outdoor) experiment for target 2

Target 2	X (cm)	Y (cm)	Z (cm)
Average	5.46	7.38	3.36
Std. Deviation	2.16	4.88	1.05
RMSE	5.81	8.66	3.50
RMSE xy	10.43		

Table 8 Evaluation of square path (outdoor) experiment for target 3

Target 3	X (cm)	Y (cm)	Z (cm)
Average	8.75	6.64	3.17
Std. Deviation	1.64	3.99	2.20
RMSE	8.89	7.61	3.78
RMSE xy	11.70		

coordinates (measured with the total station) were calculated extracting the average, the RMSE, the standard deviation for each axis (X, Y and Z) and the horizontal error RMSE xy.

As presented in Table 6 which represents the results for target 1, the average of X error is below 5 cm with a standard deviation of 3 cm while in Y axis the average error is below 8 cm with a standard deviation of 5 cm. The horizontal error is in a level of 10 cm with RMSE in 5.50 cm for the X axis and 9.50 cm for the Y axis. The results in Z axis succeed very high accuracy with 1.80 cm in average error, 1.23 cm standard deviation and a vertical error of 2.13 cm.

For target 2 as presented in Table 7, the average of X error is below 5.50 cm with a standard deviation of 2.20 cm while in Y axis the average error is below 8 cm with a standard deviation of 5 cm. The horizontal error is in a level of 10 cm with an RMSE in 6 cm for the X axis and 9.00 cm for the Y axis. The results in Z axis succeed an accuracy (vertical error) in a level of 3.50 cm with 3.40 cm in average and a standard deviation in a level of 1 cm.

Finally, for target 3 as presented in Table 8, the average of X error is below 9.00 cm with a standard deviation of 1.70 cm while in Y axis the average error is below 7 cm with a standard deviation of 4 cm. The horizontal error is in a level of 11.50 cm with an RMSE in 8.90 cm for the X axis and 7.60 cm for the Y axis. The results in Z axis succeed again

high accuracy (vertical error) in a level of 4 cm with 3.20 cm in average and a standard deviation in a level of 2.50 cm.

Initial and proposed methodology comparison

For the initial (Trigkakis et al. 2020) and proposed methodology comparison, an outdoor terrestrial experiment was conducted in multiple times maintaining the similar conditions and experiment setup. The trajectory path was in a right-angle shape while three targets were placed on the scene (Fig. 10).

As presented in the Fig. 10, the distance between the starting point of the camera and the origin marker is larger than the previous experiments of the proposed methodolgy because the monocular SLAM requires more time to initialize and start the mapping and tracking procedure.

The two approaches were tested in terms of distance estimations between the origin and target markers and the localization accuracy. The reference measurements were produced using a geodetic total station which provides an error about 1—2 cm.

The measured distances between the markers and the corresponding estimations (in average) of the initial and proposed methodology are presented in the following table:

As presented in Table 9, the distance estimations of the initial methodology is quite less accurate compared with the proposed methodology. More specifically, in target 1 the proposed methodology provides an error in a level of 3.50 cm which is almost three times smaller than the corresponding error or the initial methodology. Moreover, in targets 2 and 3, which are placed in a larger distance than the target 1, the proposed methodology maintains the error in an respectable level of about 6 cm while the initial methodology, provides an error in a level of 20 cm which is four times larger than the corresponding errors of the proposed methodology.
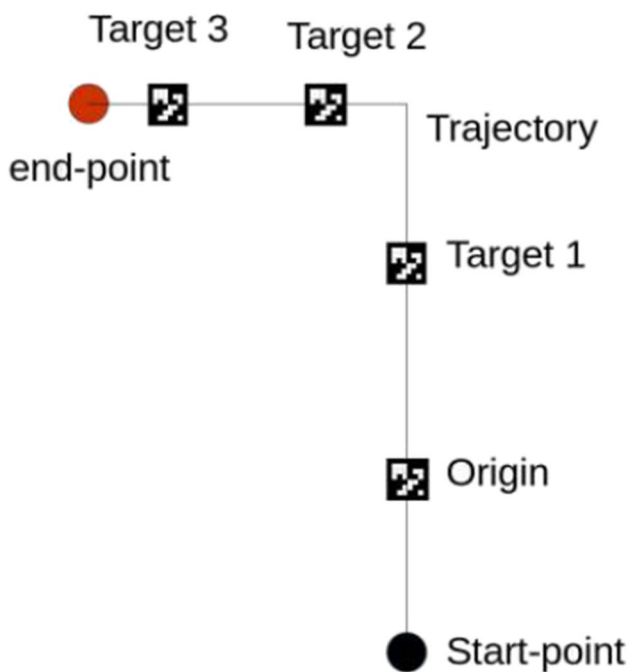
Fig. 10 Outdoor expeiment with right-angle trajectory path

The target localization results (in average) of both approaches are presented in the following table:

As observed in the Table 10, while the accuracy of the proposed methodology in Z axis is slightly higher, the horizontal accuracy is analogous with the accuracy in distance estimations. More specifically, the proposed methodology provides high accuracy in target 1 in a level of 3 cm which is quite close to the total station's accuracy while the initial approach produces an error in a level of 8.50 cm. Moreover, in targets 2 and 3, the proposed methodology provides an error of about 6 and 7.50 cm respectively while the initial appoach localizes the targets with an error about 19.50 cm which four times higher than the proposed metdhology.

In short, the proposed methodology proved its superiority compared with the initial approach (Trigkakis et al. 2020) while its estimations in distances about 6 m or below are close to the accuracy of land surveying which is until now the most accurate and reliable method for point localization in applied Geodesy. Thus, the results of the proposed methodology are considered as sufficient since the target localization is povided using a software-based approach while the land surveying utilizes the mechanical parts of total station in order to measure distances and angles aiming to produce the corresponding target localization.
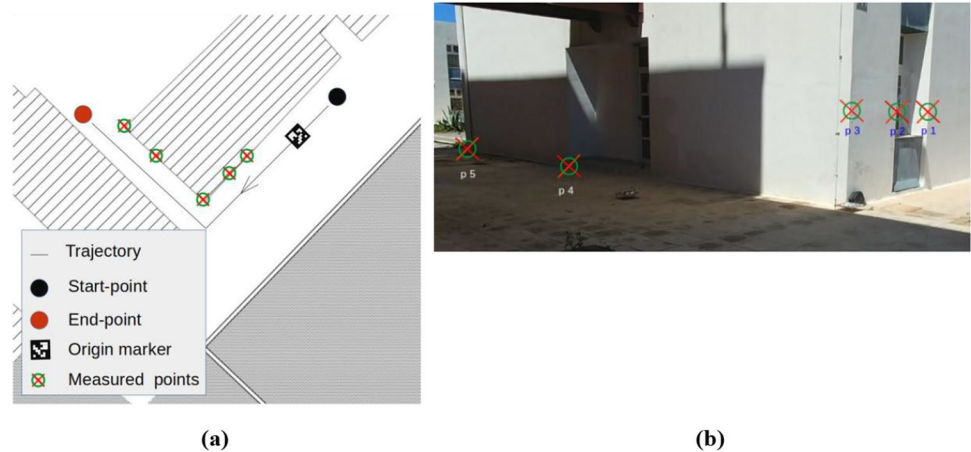
Validation of characteristic points

This experiment was performed in order to validate the coordinates of characteristic points from the extracted point cloud. Characteristic points at the sense of topography are points that compose significant geometric features in a surveying area e.g. building corners, plot boundaries etc. The purpose of the experiment was to find the accuracy of five characteristic points in two sides of a building (Fig. 11). The first three points (point 1, point 2 and point 3) are building corners with about 1.5 m height from the ground while the final two points (point 4 and point 5) building corners in the ground. The characteristic points were measured with the total station in order to calculate the ground truth of the points. The results are presented in Table 11.

As presented in Table 11, the horizontal error in point 1, point 2, point 4 and point 5 is below 10 cm while the vertical error in point 1, point 3, point 4 and point 5 is below 5 cm. It's worth to observed that point 4 have horizontal error 0.50 cm and vertical error 1.50 cm. The points are mainly created due to intensive contrast and texture features in the scene in order to reconstruct the geometry of a 3D environment. Thus, points 4 and 5 which are defined in building corners on the ground, succeed refined results, compared to points 1, 2 and 3 which were defined + 1.50 m above the ground. Points 1 and 2, defined in building corners with intensive color contrast have greater horizontal accuracy than point 3, while the high vertical accuracy of point 3 (0.71 cm) is due to the rotation of the camera direction in the horizontal plane (in order to follow the right angle shape

Table 9 Evaluation in distance estimations of the initial (I.M) and proposed (P.M) methodology

Target markers	Distances (I.M)	Distances (P.M)
Origin—Target 1 (Ground Truth)	340.5	340.5
Origin—Target 1 (Estimation)	348.98	337.19
Error (cm)	8.48	3.31
Origin -Target 2 (Ground Truth)	696.21	696.21
Origin -Target 2 (Estimation)	715.41	691.16
Error (cm)	19.20	5.05
Origin—Target 3 (Ground Truth)	722.46	722.46
Origin—Target 3 (Estimation)	742.39	715.94
Error	19.93	6.52

Table 10 Evaluation of target localization of the initial (I.M) and proposed (P.M) methodology

(cm)	Initial approach		Proposed methodology	
	X Y error	Z error	X Y error	Z error
Target 1	8.52	1.5	3.13	0.66
Target 2	19.22	6.5	5.85	5.86
Target 3	19.81	7.75	6.78	7.53

Fig. 11 a A map with the characteristic measured points (p1, p2, p3, p4, p5) and path trajectory **b** The study area in which the characteristic points are depicted

(a)

(b)

Table 11 Evaluation of characteristic measured points

Points	X (cm)	Y (cm)	Z (cm)	Min Dist
Point 1: G. T	98.30	325.48	144.00	9.29
Closest point	102.72	317.55	142.84	
Error	4.42	7.93	1.16	
XY error	9.08		-	
Point 2: G. T	97.18	451.63	150.0	13.90
Closest point	98.78	447.03	162.36	
Error	1.60	4.60	12.36	
XY error	4.87		-	
Point 3: G. T	93.90	600.00	149.00	13.01
Closest point	102.06	589.58	148.29	
Error	8.16	10.42	0.71	
XY error	13.23		-	
Point 4: G. T	953.85	613.26	-7.00	1.5802
Closest point	954.00	612.82	-8.51	
Error	0.15	0.44	1.51	
XY error	0.46		-	
Point 5: G. T	1699.7	619.38	-10.00	8.64
Closest point	1698.75	626.35	-15.00	
Error	0.95	6.97	5.00	
XY error	7.03		-	

of the trajectory, Fig. 11) allowing a wider stereo vision information for point 3.

Discussion

The proposed methodology encounters the issue of high accuracy mapping with the use of stereo vision. The methodology estimates the marker and arbitrary point locations in 3D space of an unknown environment using a visual marker and a stereo camera with or without a UAV. At first, the stereo camera is guided through a desired path, identifies

the markers and maps the surroundings. In a second step, a local coordinate system is created for the scene with an origin defined by the initial marker that the camera comes across, while the coordinates of the target markers and the arbitrary points of the point cloud are calculated based on the origin marker. In order to evaluate the methodology, different sets of experiments were performed in terms of study area (indoor / outdoor), the use or not of a UAV, the number and location of markers, the trajectory paths and the point cloud validation, while in the last UAV experiment, a statistical evaluation was conducted based on 50 separate records of the scene. Although the methodology's accuracy is highly correlated with some factors that will be further mentioned, it achieved quite remarkable horizontal and vertical accuracy considering the minimal equipment used.

The experiments separated in two main categories: non-UAV and UAV experiments. Concerning the non-UAV experiments, the maximum horizontal error is in a level of 6 cm and the minimum error, in a level of 2 cm for indoor environment while in outdoor environment the maximum horizontal error is less than 9 cm and the minimum error less than 2 cm. It is observed that in non-UAV experiments, the horizontal error is linearly correlated with the distance between the target markers and the origin marker and the horizontal error is increasing in proportion with the distance (Table 12).

It's worth to mention that the outdoor non-UAV experiment succeeds higher horizontal accuracy than the indoor experiment. For instance, as referred in Table 12, the horizontal error in target 2 of outdoor pi-shaped experiment is 3.42 cm in a distance of 9.26 m from the origin while in the indoor experiments the error in target 2 is in a level of 5.50 cm in a distance of 7 m. Moreover, the error in target 3 of outdoor experiment is 8.81 cm in a distance of 17.33 m, only 3 cm higher than the corresponding errors of indoor experiments (5.99, 5.33 cm) which have a 10 m shorter distance (7.22 m). This difference of horizontal error in indoor

Table 12 Distance (from the origin to target marker) and horizontal error in non-UAV experiments

Experiment	Right angle path (indoor)		Step-shaped path (indoor)		Pi-shaped path (outdoor)	
	Distance	Hor. error	Distance	Hor. error	Distance	Hor. error
Target 1	3.4 m	2.57 cm	3.40 m	2.08 cm	2.44 m	1.7 cm
Target 2	6.96 m	5.30 cm	6.96 m	4.49 cm	9.26 m	3.42 cm
Target 3	7.22 m	5.99 cm	7.22 cm	5.33 cm	17.33 m	8.81 cm

Table 13 Distance (from the origin to target marker) and vertical error in non-UAV experiments

Experiment	Right angle path (indoor)		Step-shaped path (indoor)		Pi-shaped path (outdoor)	
	Distance	Ver. error	Distance	Ver. error	Distance	Ver. error
Target 1	3.4 m	3.10 cm	3.40 m	0.08 cm	2.44 m	7.80 cm
Target 2	6.96 m	9.06 cm	6.96 m	5.14 cm	9.26 m	2.50 cm
Target 3	7.22 m	8.78 cm	7.22 cm	7.31 cm	17.33 m	24.80 cm

and outdoor experiments is due to the refined information that the sensors acquire in outdoor environments because of physical illumination while the geometry of the outdoor scene such as buildings and trees aids the methodology to provide more accurate results for target markers that are located close to features like a building.

Regarding the vertical error, the distance doesn't seem to affect the results in general as presented in Table 13 below.

The main reason that decreases the vertical accuracy in non-UAV experiments is the variable height of the camera because of the physical vertical motion of the human body while walking and the different altitude levels of the ground. Thus, in indoor experiments where the only vertical motion of the camera is due to walking, the maximum vertical error is less than 9.50 cm and the minimum vertical error is less than 1 cm. On the other hand, the combination of these two factors in outdoor environment is able to highly decrease the vertical accuracy. As observed in Table 11, while the vertical error of targets 1 and 2 is less than 8 cm the error of target 3 reaches in a level of 25 cm. The overall vertical accuracy can be refined if the camera height is stable as proved in the UAV-experiments. Also a gimbal system on the terrestrial, non-UAV measurements could further compensate possible vertical errors.

In the UAV experiments, the stereo camera is attached on the custom-made UAV and interfaces with the embedded computing system while the flight height is about 3 m. The first UAV experiment was performed in a straight path trajectory and succeeded horizontal error less than 7.50 cm and vertical error less than 1.5 cm. Even though from the first experiment it is evident that the horizontal accuracy is quite promising and the vertical error seems to be decreased compared with the non-UAV methodology, it was difficult to evaluate the UAV-enabled methodology because the wind during flight and the drifts of UAV from the trajectory path had significant impact to the results. Thus, a square path experiment was conducted in which the UAV performed 50 flights in a time-step of several days in sunny weather and

Table 14 RMSExy and RMSEz per target

Target 1 (cm)		Target 2 (cm)		Target 3 (cm)	
RMSE xy	RMSE z	RMSE xy	RMSE z	RMSE xy	RMSE z
10.59	2.13	10.43	3.50	11.70	3.78

low wind speed conditions in order to evaluate the UAV-enabled accuracy statistically. The square shape of the path was chosen because of its simplicity in order to interpret the results and simulate a real-world survey area.

As presented in Table 14, the horizontal error in all targets is less than 12 cm while as mentioned all RMSE errors in X and Y axes are less than 9 cm (Tables 6, 7 and 8). The standard deviation of X axis varies in a range of 1.50 to 3 cm while in Y axis, reaches a level of 5 cm (Table 15). This variability in horizontal error is highly correlated with the UAV drifts from the trajectory path during the flight.

Most of the flights from this experiment preserved a stable and accurate square trajectory path which produced refined results as observed in Tables 16, 17 and Fig. 12. The horizontal error in target 1 and target 3 is in a level of 8 cm with a standard deviation less than 3.10 cm in X axis and less than 2.90 cm in Y axis, while in target 3 the error is in a level of 11 cm with a standard deviation 1.63 cm in X and 2.63 in Y axis.

In Fig. 12, is visualized the impact of the UAV drifts in the results. In target 3 the accrracy is about 1 cm higher while in targets 1 and 2, the error is decreased with 2 and 3 cm respectively in the results without a drift. The figure above and Tables 12 and 14 determines that the weather conditions and more specifically the wind speed which is the most important factor for a stable flight are quite important for the effectiveness of the methodology. It's worth to mention that the horizontal errors are very similar with the non-UAV experiments which prove that the flight height of 3–4 m has no impact to the results.

Table 15 Standard deviation for X, Y and Z axes per target

Target 1 (cm)			Target 2 (cm)			Target 3 (cm)		
Std D. X	Std D. Y	Std D. Z	Std D. X	Std D. Y	Std D. Z	Std D. X	Std D. Y	Std D. Z
2.93	4.89	1.23	2.1	4.88	1.05	1.64	3.99	2.20

Table 16 RMSExy and RMSEz per target for the flights without UAV drifts

Target 1 (cm)		Target 2 (cm)		Target 3 (cm)	
RMSE xy	RMSE z	RMSE xy	RMSE z	RMSE xy	RMSE z
7.51	2.43	8.21	3.45	10.60	4.47

On the other hand, the vertical error in terms of RMSE derived from all the records (drift and without drift) is less than 2.20 cm in target 1 with standard deviation of 1.23 cm, while for targets 2 and 3 the vertical errors are less than 3.80 cm with standard deviation of 1.05 and 2.20 respectively. Comparing with non-UAV experiments the vertical error is significantly lower, which verifies that a stable height of the camera is highly recommended for the improvement of vertical accuracy.

Concerning the point cloud that the methodology exports, although it cannot be used for target estimations in specific points, the experiment in Section 3.3 proves that most of the characteristic points that were measured with the total station were identified, generating a denser point cloud close to the measured points because of the color and depth contrast of features. Thus, the point cloud constructs the 3D environment with a centimeter-level of accuracy which can be visualized and measured through the visualization module.

To sum up, as the results proved, the proposed methodology is able to be used as a surveying alternative in outdoor or indoor environment with or without a UAV. More specifically, two real-case scenarios of mapping can be covered: the localization of points in an accessible urban or vegetated area and the mapping of a rocky or hard to walk area. Regarding the urban or vegetated area, the user is able to walk along the trajectory path with the stereo-camera in a steady motion, achieving a 3D error (XY and Z error) in a level of 6 cm in a distance of about 20 cm between the origin and target markers.. On the other hand, in a hard-to walk area, the user is able to place the visual markers in the desired locations and then, using a UAV with an attached stereo camera to map the area and estimate the target coordinates, maintaining a steady motion for the camera. The 3D error according to the evaluation above is in a level of 6.11 cm and 7 cm with and without drifts respectively. Thus, with or without the use of a UAV, the methodology covers the most real-life cases in terms of the environment variability.

Table 17 Standard deviation for X, Y and Z axes per target for the flights without UAV drifts

Target 1 (cm)			Target 2 (cm)			Target 3 (cm)		
Std D. X	Std D. Y	Std D. Z	Std D. X	Std D. Y	Std D. Z	Std D. X	Std D. Y	Std D. Z
3.04	1.72	1.62	2.48	2.84	1.26	1.63	2.63	2.30

Fig. 12 Horizontal error in terms of RMSE of flights with drift and flights with no drift from trajectory path
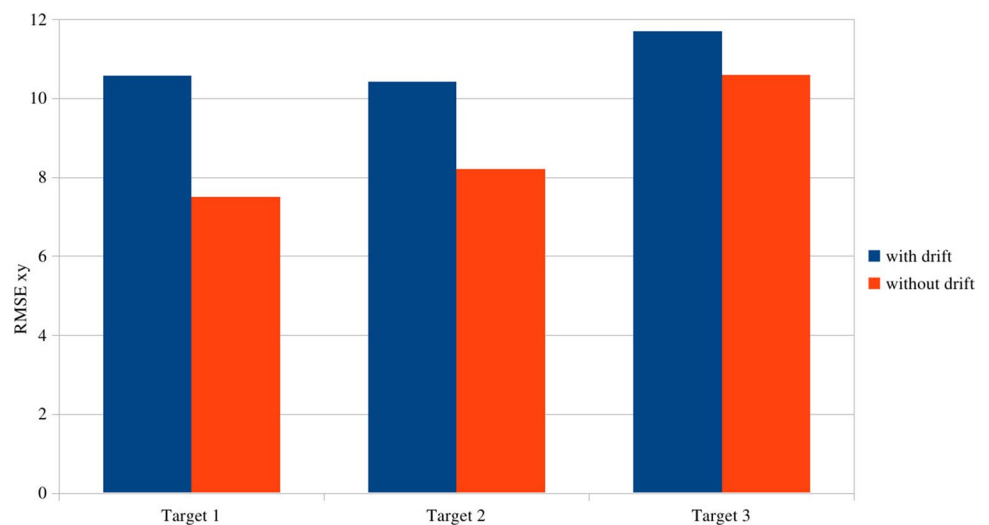
For the localization of the targets with high performance as presented in the results section, a certain way of mapping have to be followed in order the methodology to provide qualitative results. The camera has to follow a trajectory that begins a few meters before the origin marker and then proceed, approaching close and overtaking all the markers (origin and targets) maintaining a non-complicated path. This technique, provides a reasonable sequence of frames to the system, aiding the feature extraction process, camera pose estimations and consequently marker localization using the M.L.C and P.A methods. The sudden unreasonable camera movements or a complicated camera trajectory where the camera doesn't approach the markers directly maintaining a steady direction, will significantly decrease the accuracy.

Moreover, the large distance (about 20 m or above) between the origin and target markers can decrease the accuracy of the methodology especially when the origin and the target markers are not colinear. However, the error of each target is dependent only on the location of the origin marker and not affected by the other markers. For instance, in the UAV-based experiment of the square path, the error of the target 1, target 2 and target 3 is not accumulated, instead each of these targets is associated only with the origin marker. Nevertheless, the increased distance between the origin and target markers can affect the accuracy.. This issue is a limitation of the methodology which can be encountered with the integration of a life-long SLAM architecture (as referred by the computer vision community) which constitutes one of the main future goals of the proposed methodology.

Comparing with the monocular approach of this methodology (Trigkakis et al. 2020), the stereo approach succeeded highly improved results in a more complex experimentation. In Trigkakis et al. (2020) the RMSExy for the target marker is 41 cm and RMSEz is 6.4 cm while in the stereo approach the horizontal accuracy reaches a level of 11 cm and about 3 cm of vertical accuracy. The accuracy in stereo approach is about four times better in terms of RMSExy and two times better in terms of RMSEz proving that the stereo approach produces more accurate and sophisticated results.

Conclusions

The present study, proposes an alternative mapping methodology which focuses on surveying in indoor and outdoor environments. The main contribution of this study is that solves the issue of precise localization of visual points in a 10 cm of accuracy using only a stereo camera, a conventional computing system and a visual marker while the use of a UAV is integrated and tested.

In other words, instead of similar computer vision systems, this study focuses on the localization with high accuracy using a coordinate system defined in the scene, based on the pose of a physical marker. This fact, makes the methodology completely comparable with the traditional surveying process in which the measurements are conducted using a geodetic total station and the coordinate system is defined in the scene, using the internal geometry of the total station on a reference point. However, while the traditional surveying requires significant human effort and a quite costly equipment, the proposed methodology is conducted with just a walk in a specific path of the scene, being able to provide the coordinate estimations in a few minutes.

Moreover, the two UAV and non-UAV approaches of the methodology provide flexibility to the users in terms of the area morphology, the type of environment and the desired accuracy. For instance, for the mapping of a steep and rocky survey area, the UAV approach is recommended, while for an indoor environment the non-UAV use can achieve respectable accuracy. For a surveying project which requires high vertical accuracy, the UAV approach will produce refined results while for the mapping of a plot, the non-UAV approach can achieve high horizontal accuracy.

Although the present study provides significant advantages compared with the traditional surveying, it contains limitations that are still under research. Although the mapping process in the field is quite straightforward, is affected from factors such as the wind speed, the variable height of the camera, the illumination conditions and the distance from the origin which decrease the accuracy of the system. Also, even in an ideal scenario where the accuracy is able to decrease under the level of 10 cm, is still far away from the accuracy of traditional surveying which achieves an accuracy level of 1—2 cm.

However, knowing the factors that decrease the accuracy, is setting the basis for the future research of the optimization of the methodology. For instance, a smaller and more stable UAV could perform more accurate flights with even higher horizontal accuracy or a gimbal device could stabilize the camera during data collection. Moreover, the optimization of the SLAM algorithm in order to improve its robustness in different illumination conditions or to maintain its accuracy in larger distances, could increase even more the effectiveness of this study.

Taking in to account all the above constraints, the main goal of this study for the future is to achieve accuracy in a level of 2–5 cm. This level of accuracy in combination with the minimum cost of equipment is possible to change the way of mainstream surveying is conducted and add a new simple and cost-effective surveying technique in terms of scientific methodology and equipment.

Declarations

Competing interests The authors declare no competing interests.

References

Ampadu EG, Gebreslasie M, Ponce AM (2020) Mapping natural forest cover using satellite imagery of Nkandla forest reserve, KwaZulu-Natal, South Africa. Remote Sens Appl: Soc Environ 18. https://doi.org/10.1016/j.rsase.2020.100302

Animesh S, Harsh S, Mangal K (2019) Autonomous detection and tracking of a high-speed ground vehicle using a quadrotor UAV. In: Proceedings of the AIAA Scitech Forum, San Diego, California, 7–11 January

Bañón L, Pagán JI, López I, Banon C, Aragonés L (2019) Validating UAS-based photogrammetry with traditional topographic methods for surveying dune ecosystems in the Spanish Mediterranean coast. J Mar Sci Eng 7:297. https://doi.org/10.3390/jmse7090297

Bobbe M, Kern A, Khedar Y, Batzdorfer S, Bestmann U (2017) An automated rapid mapping solution based on ORB SLAM 2 and Agisoft Photoscan API. In: Proceedings of the IMAV, Toulouse, France

Burri M, Nikolic J, Gohl P, Schneider T, Rehder J, Omari S, Achtelik M, Siegwart R (2016) The euroc micro aerial vehicle datasets. Int J Robot Res. https://doi.org/10.1177/0278364915620033

Carrera-Hernández JJ, Levresse G, Lacan P (2020) Is UAV-SfM surveying ready to replace traditional surveying techniques? Int J Remote Sens 41(12):4820–4837. https://doi.org/10.1080/01431161.2020.1727049

Chiang KW, Tsai GJ, Chang HW, Joly C, EI-Sheimy N (2019) Seamless navigation and mapping using an INS/GNSS/grid-based SLAM semi-tightly coupled integration scheme. Inf Fusion 50:181–196. https://doi.org/10.1016/j.inffus.2019.01.004

Chuang R (2018) Mapping surface breakages of the 2018 Hualien earthquake by using UAS photogrammetry. Terr Atmospheric Ocean Sci 30:351–366. https://doi.org/10.3319/TAO.2018.12.09.02

Eldén L (1982) A weighted pseudoinverse, generalized singular values, and constrained least squares problems. BIT 22:487–502. https://doi.org/10.1007/BF01934412

Gabrlik P (2015) The use of direct georeferencing in aerial photogrammetry with micro UAV. IFAC-PapersOnLine 48:380–385. https://doi.org/10.1016/j.ifacol.2015.07.064

Gao X, Wang R, Demmel N, Cremers D (2018) LDSO: direct sparse odometry with loop closure. arXiv

Garrido-Jurado S, Muñoz SR, Madrid-Cuevas FJ, Medina-Carnicer R (2016) Generation of fiducial marker dictionaries using mixed integer linear programming. Pattern Recogn 51:481–491. https://doi.org/10.1016/j.patcog.2015.09.023

Geiger A, Ziegler J, Stiller C (2011) Stereoscan: dense 3d reconstruction in real-time, in Intelligent Vehicles Symposium (IV), pp. 963–968. https://doi.org/10.1109/IVS.2011.5940405

Geiger A, Lenz P, Urtasun R (2012) Are we ready for autonomous driving? The kitti vision benchmark suite. In: Conference on CVPR

Geneva P, Eckenhoff K, Lee W, Yang Y, Huang G (2020) Openvins: a research platform for visual-inertial estimation. In: 2020 ICRA, pp 4666–4672. https://doi.org/10.1109/ICRA40945.2020.9196524

Giubilato R, Chiodini S, Pertile M, Debei S (2018) An experimental comparison of ROS-compatible stereo visual SLAM methods for planetary rovers. https://doi.org/10.1109/MetroAeroSpace.2018.8453534

Haque A, Elsaharti A, Elderini T, Elsaharty MA, Neubert J (2020) UAV autonomous localization using macro-features matching with a CAD model. Sensors 20:743. https://doi.org/10.3390/s20030743

Helgesen HH, Leira FS, Bryne TH, Albrektsen SM, Johansen TA (2019) Real-time georeferencing of thermal images using small fixed-wing UAVs in maritime environments. ISPRS J Photogramm Remote Sens 154:84–97. https://doi.org/10.1016/j.isprsjprs.2019.05.009

Jende P, Nex F, Gerke M, Vosselman G (2018) A fully automatic approach to register mobile mapping and airborne imagery to support the correction of platform trajectories in GNSS-denied urban areas. ISPRS J Photogramm Remote Sens 141:86–99. https://doi.org/10.1016/j.isprsjprs.2018.04.017

Kalacska M, Lucanus O, Arroyo-Mora JP, Laliberté É, Elmer K, Leblanc G, Groves A (2020) Accuracy of 3D landscape reconstruction without ground control points using different UAS platforms. Drones 4:13. https://doi.org/10.3390/drones4020013

Kuroswiski AR, Oliveira NMF, Shiguemori EH (2018) Autonomous long-range navigation in GNSS-denied environment with low-cost UAV platform. In: Proceedings of the SysCon, Vancouver, Bc, Canada, 23 - 26 April. https://doi.org/10.1109/SYSCON.2018.8369592

Labb M, Michaud F (2013) Appearance-based loop closure detection for online large-scale and long-term operation. IEEE Trans Robot 29(3):734–745. https://doi.org/10.1109/TRO.2013.2242375

Lassila MM (2018) Mapping mineral resources in a living land: Sami mining resistance in Ohcejohka, northern Finland. Geoforum 96:1–9. https://doi.org/10.1016/j.geoforum.2018.07.004

Li J, Besada JA, Bernardos AM, Tarrío P, Casar JR (2017) A novel system for object pose estimation using fused vision and inertial data. Inf Fusion 33:15–28. https://doi.org/10.1016/j.inffus.2016.04.006

Ma Y, Li Q, Chu L, Zhou Y, Xu C (2021) Real-time detection and spatial localization of insulators for UAV inspection based on binocular stereo vision. Remote Sens 13:230. https://doi.org/10.3390/rs13020230

McRae JN, Gay CJ, Nielsen BM, Hunt AP (2019) Using an unmanned aircraft system (drone) to conduct a complex high altitude search and rescue operation: a case study. Wilderness Environ Med 30:287–290. https://doi.org/10.1016/j.wem.2019.03.004

Mishra B, Garg D, Narang P, Mishra V (2020) Drone-surveillance for search and rescue in natural disaster. Comput Commun 156:1–10. https://doi.org/10.1016/j.comcom.2020.03.012

Munguía R, Urzua S, Bolea Y, Grau A (2016) Vision-based SLAM system for unmanned aerial vehicles. Sensors 16:372. https://doi.org/10.3390/s16030372

Mur-Artal R, Tardós JD (2017) ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras. IEEE Trans Robot 33:1255–1262. https://doi.org/10.1109/TRO.2017.2705103

Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 9:62–66

Panigrahi N, Doddamani SR, Singh M, Kandulna BN (2015) A method to compute location in GNSS denied area. IEEE International CONECCT 1–5. https://doi.org/10.1109/CONECCT.2015.7383907

Partsinevelos P, Chatziparaschis D, Trigkakis D, Tripolitsiotis AA (2020) Novel UAV-assisted positioning system for GNSS-denied environments. Remote Sens 12:1080. https://doi.org/10.3390/rs12071080

Pinto AM, Matos AC (2020) MARESye: a hybrid imaging system for underwater robotic applications. Inf Fusion 55:16–29. https://doi.org/10.1016/j.inffus.2019.07.014

Pire T, Fischer T, Civera J, Cristforis P, Berlles J (2015) Stereo parallel tracking and mapping for robot localization in Proc. IROS, pp 1373–1378. https://doi.org/10.1109/IROS.2015.7353546

Qin T, Cao S, Pan J, Shen S (2019) A general optimization-based framework for global pose estimation with multiple sensors, arXiv

Romero-Ramirez FJ, Muñoz-Salinas R, Medina-Carnicer R (2018) Speeded up detection of squared fiducial markers. Image Vis Comput 76:38–47. https://doi.org/10.1016/j.imavis.2018.05.004

Rosinol A, Abate M, Chang Y, Carlone L (2020) Kimera: an open-source library for real-time metric-semantic localization and mapping. In: IEEE Intl. Conf. on Robotics and Automation (ICRA)

Rublee E, Rabaud V, Konolige K. and Bradski G, (2011) ORB: an efficient alternative to SIFT or SURF. International conference on computer vision, pp 2564–2571. https://doi.org/10.1109/ICCV.2011.6126544

Sahoo B, Biglarbegian M, Melek W (2021) Monocular visual inertial direct SLAM with robust scale estimation for ground robots/vehicles. Robotics 10:23. https://doi.org/10.3390/robotics10010023

Samuel B (2004) Introduction to inverse kinematics with Jacobian transpose, pseudoinverse and damped least squares methods. IEEE Trans Robot Autom 17:16

Schneider T, Dymczyk M, Fehr M, Egger K, Lynen S, Gilitschenski I, Siegwart R (2018) Maplab: an open framework for research in visual-inertial mapping and localization. IEEE Robot Autom Lett. https://doi.org/10.1109/LRA.2018.2800113

Sharafutdinov D, Griguletskii M, Kopanev P, Kurenkov M, Ferrer G, Burkov A, Gonnochenko A, Tsetserukou D (2021) Comparison of modern open-source visual SLAM approaches, arXiv

Shi X, Li D, Zhao P, Tian Q, Tian Y, Long Q, Zhu C, Song J, Qiao F, Song L, Guo Y, Wang Z, Zhang Y, Qin B, Yang W, Wang F, Chan R, She Q (2020) Are we ready for service robots? The OpenLORIS-Scene datasets for lifelong SLAM. ICRA 2020, pp 3139–3145

Sumikura S, Shibuya M, Sakurada K (2019) Open-VSLAM: a versatile visual SLAM framework. In: Proceedings of the 27th ACM International Conference on Multimedia, MM '19, New York, NY, USA, pp 2292–2295. https://doi.org/10.1145/3343031.3350539

Tang J, Chen Y, Niu X, Wang L, Chen L, Liu J, Shi C, Hyyppä J (2015) LiDAR scan matching aided inertial navigation system in GNSS-denied environments. Sensors 15:16710–16728. https://doi.org/10.3390/s150716710

Tomaštík J, Mokroš M, Saloň Š, Chudý F, Tunák D (2017) Accuracy of photogrammetric UAV-based point clouds under conditions of partially-open forest canopy. Forests 8:151. https://doi.org/10.3390/f8050151

Trigkakis D, Petrakis G, Tripolitsiotis A, Partsinevelos P (2020) Automated geolocation in urban environments using a simple camera-equipped unmanned aerial vehicle: a rapid mapping surveying alternative? ISPRS Int J Geo-Inf 9:425. https://doi.org/10.3390/ijgi9070425

Urzua S, Munguia R, Grau A (2017) Vision-based SLAM system for MAVs in GPS-denied environments. Int J Micro Air Veh 9:283–296. https://doi.org/10.1177/1756829317705325

Usenko V, Demmel N, Schubert D, Stückler J, Cremers D (2019) Visual-inertial mapping with non-linear factor recovery. arXiv. https://doi.org/10.1109/LRA.2019.2961227

Vrba M, Saska M (2020) Marker-less micro aerial vehicle detection and localization using convolutional neural networks. IEEE Robot Autom Lett 5:2459–2466. https://doi.org/10.1109/LRA.2020.2972819

Vrba M, Heřt D, Saska M (2019) Onboard marker-less detection and localization of non-cooperating drones for their safe interception by an autonomous aerial system. IEEE Robot Autom Lett 4:3402–3409. https://doi.org/10.1109/LRA.2019.2927130

Xu L, Feng C, Kamat VR, Menassa CC (2019) An Occupancy Grid Mapping enhanced visual SLAM for real-time locating applications in indoor GPS denied environments. Autom Constr 104:230–245. https://doi.org/10.1016/j.autcon.2019.04.011

Yang T, Ren Q, Zhang F, Xie B, Ren H, Li J, Zhang Y (2018) Hybrid camera array-based UAV auto-landing on moving UGV in GPS-denied environment. Remote Sens 10:1829. https://doi.org/10.3390/rs10111829

Zahran S, Moussa A, El-Sheimy N (2018) Enhanced UAV navigation in GNSS denied environment using repeated dynamics pattern recognition. IEEE/ION PLANS, 1135–1142. https://doi.org/10.1109/PLANS.2018.8373497

Zhang C, He T, Zhan Q, Hu X (2019) Visual navigation based on stereo camera for water conservancy UAVs. In: Proceedings of the ICIST, Hulunbuir, China.https://doi.org/10.1109/ICIST.2019.8836851