Evaluation of Visual Localization Systems in Underground Mining

Christoforos Kanellakis and George Nikolakopoulos

Abstract— In this article an evaluation of current technology on visual localization systems for underground mining is presented. The proposed study is considered to be the first step among others towards enabling vision-based mine inspection using Unmanned Micro Aerial Vehicles (UAVs). Furthermore, the aim of this article, is to verify applicable and reliable low cost existing methods and technologies for the problem of UAV localization in harsh and challenging environments. More specifically field trials were performed in one of the biggest mines in Europe, the iron ore mine of LKAB at Kiruna in Sweden. In this experimental evaluation the sensors employed were a RGB-D camera (Kinect for Windows) and a web camera (Playstation 3 Eye) in two configurations, as a stereo rig and as a monocular visual sensor. The processing of the stored data from the experiments will provide an insight into the efficiency of these sensors. Additionally, will identify what further technological and research developments are required to develop affordable autonomous UAV solutions for improving the underground mining production tasks.

Fig. 1: Types of Unmanned Aerial Vehicles

I. INTRODUCTION

Unmanned Aerial Vehicles are not a new concept since they already exist for more than a decade. They consist a major field of research that is continuously growing and attracting many resources. This massive interest has led to the development of various aircraft types from actual size helicopters, fixed wing planes to micro rotor-crafts, as depicted in Figure 1 with: a) [1], b) [2], c) [3], d) [4], aiming to operate in different environments [5]. Nowadays the main effort of scientific community is focused on the field of Micro UAVs.

Micro UAVs are agile platforms able to explore an area in various speeds (low or high), to hover over a target and perform maneuvers in close distances [6]. Recently, mining industry showed interest to invest in Micro UAV applications. These platforms could be recruited in the production line of a mine contributing to different demands of production processes [7], while increasing automation, safety and operational efficiency. Briefly, aerial vehicles can be semi or fully autonomous operated to acquire and process 3D spatial data in an underground area that is difficult to access. Moreover, a UAV can explore an area prior to

sending in production machines or humans and plan for the next step in the process. Some of the possible real time applications for this kind of technology is visual inspection of collapsed or active areas, old mines inspection, condition monitoring, information collection for rescue missions and inspection or supervision of other machines while working. The conditions in such environments are harsh with limited light and technical infrastructure, as it can be depicted in Figure 2 and therefore the mining versions of MAVs should be robust and cost effective against those factors.

Fig. 2: Loading in a production area. Photograph by Fredrik Alm

In these operational scenarios, one of the major tasks provided by UAVs is the 3D Mapping and Localization for path planning and collision free navigation for long term and

large scale operations. At this point is should be highlighted that to the authors best knowledge, there has not been a significant amount of references in the related literature so far regarding the specific topic of underground localization. In [8], a UAV system for the environment exploration and mapping, by fusing ultrasonic and camera sensors has been developed. In the presented algorithm, the 2D marker planar data, extracted from the image and the depth measurements, from the ultrasonic sensor, were merged and computed the UAVs position, while other ultrasonic sensors were detecting the obstacles. In the sequel, this information was further processed in order to build a map of the surrounding area. In the presented evaluation scenario, it was assumed that the quadrotor was able to move vertically up and down, without rotating around its axis. In [9] a low cost quadrotor being capable of visual navigation in unstructured environments by using on board processing has been developed. The main components of this work have been a SLAM system, an extended Kalman Filter and a PID controller. This research approach proposed a novel closed-form maximum likelihood estimator to remove the measurement noise and recover the absolute scale of the visual map. In [10], the hardware, software and mechanical parts of an aerial platform that performs autonomous exploration in cluttered environments has been developed. The platform was equipped solely with a stereo camera and an inertial sensor. The proposed visual inertial navigation framework fused data from a camera and an IMU to implement position tracking control of the platform and it was robust against visual odometry failures. In this case, the UAV was controlled to follow obstacle free paths, while the system was experimentally evaluated in real scale in a tunnel environment, simulating a coal mine (some simplifications like lightning conditions were part of the experiment). Finally, in [11] a first evaluation step in understanding underground localization applications has been presented, that was more realistic than [10]. More specifically, a hexacopter equipped with a Visual Inertial sensor and a laser scanner was manually guided across a vertical mine shaft to collect data for post-processing. The extracted information from the measurements have been utilized to create a 3D mesh of the environment and localize the vehicle.

Although, the presented technology is still in an early stage and encounters serious challenges, it has been already depicted that autonomous UAVs in underground mines are an emerging area. The novelty and scientific contribution of this article is divided in three parts. Firstly, focuses in the application of low cost existing vision based localization techniques and visual sensors in underground environments. Additionally, presents their experimental evaluation under difficult conditions as a future guideline to enable further developments in environmental perception and localization in mine inspection. A second contribution of this article is the fact that such experimental results, from real life harsh environments in underground mines and especially in real production areas, have never been reported before. Moreover, the experimental evaluation of the utilized vision algorithms, as well as their performance analysis have never been appeared before in the scientific literature. Finally, the third contribution of this article is the identification of specific research directions and challenges for improving the performance and the utilization of visual localization systems in underground mining.

To fulfill these contributions, experiments were performed with 3 different visual sensors, a RGB-D, a stereo camera configuration and a monocular camera, in a production area tunnel of the LKAB's iron ore mine at Kiruna in Sweden. In the performed experiments, each sensor was examined separately in order to identify the advantages and disadvantages and create a reference for further modifications and developments. Since all the sensors were based in RGB modules, visual SLAM techniques were employed in this work for the localization and mapping tasks.

The rest of this article is structured as it follows. In Section II the SLAM methods that have been integrated and utilized are described, followed by the experiment design in section III and the corresponding experimental results in Section IV. In V an extended discussion regarding the further developments in the field is being presented, while conclusions are drawn in Section VI.

II. METHODS

As it has been stated in Section I, the aim of this article, is to evaluate the performance of the visual localization systems in underground mining. Therefore, the employed methods to be evaluated should be able to perform "online" according to vehicle dynamics and large scale operations. In this way, establishing long term localization while meeting specific time constraints could guarantee stability for the aerial platform. Regarding the camera sensors, visual SLAM [12] approaches have been utilized to post process the data collected from the mine experiments. The purpose of the cameras is two folded, firstly they are utilized for the ego-motion estimation [13] of the sensor, providing information of the pose with respect to the starting point. Secondly using the estimated pose information point clouds are sequentially stitched in order to reconstruct the surrounding environment [14].

In general, there have been two kinds of visual sensors that were employed in these experiments: a) those that are able to calculate depth in absolute scale, Kinect 2 and the stereo camera rig, and b) a single camera sensor, consisting of a monocular vision system that it is able to perceive depth from the motion up to a scale. In the performed experiments, the technique used for the depth sensors has been the RTAB-Map SLAM [15], an appearance based Localization and Mapping algorithm, which can approach the real time performance. More specifically, RTAB Map is a SLAM approach that consists of three parts, the visual odometry and the loop closure detection, as well as the graph optimization part. The main characteristics of the former part are the constant time complexity and the update rate around 20 Hz. During pose estimation, 2-diamensional visual feature extraction is

<div align="center">

(a) PS3 EYE cameras   (b) Kinect for Windows

Fig. 3: Visual Sensors

</div>

performed to find distinctive points in surrounding and derive their depth. Afterwards, correspondences between identified features are calculated from consecutive camera frames. Finally, from corresponding features with known 3D positions, a RANSAC [16] rigid transformation estimation is performed to recover the ego-motion of the camera. The mapping is calculated in slower rates like 1Hz. In parallel loop closure between the current frame with all the stored frames in a graph is continuously checked. When a loop closure is detected new constraints are added to the graph optimization minimizing errors of the estimated poses. Finally, each node of the graph contains a dense point cloud of the scene which is aligned with its neighbors depending on the node pose. The aforementioned approach implements specific memory management scheme for online usage.

For the monocular camera based localization, ORB-SLAM [17] technique has been evaluated. The ORB SLAM is a feature-based approach that combines ORB feature extraction with PTAM [18] and thus making the overall performance of the algorithm more robust and simpler. This system extracts and employs the same features for a series of tasks like tracking, mapping, relocalization and loop closure. The tracking part detects and maintains camera pose information, the local mapping part is employed to guarantee optimal reconstruction of camera surroundings and is constantly searching for new correspondences. Finally loop closure is used to compensate for accumulating drift in the system. On top of that, the flexible planar or non-planar stereo initialization and the Bag of Words [19] loop detection and closure (Essential Graph) techniques, make the ORB SLAM a powerful tool for monocular localization and mapping. It should be highlighted that tracking mapping and loop closure tasks consist separate threads that run in parallel.

Regarding all the performed experiments, the data collection and the localization methods were implemented using the Robotic Operation System (ROS) [20] framework.

III. EQUIPMENT AND EXPERIMENTS

Underground localization and inspection are the two main objectives studied with a specific focus on cost-effective and reliable solutions. To this end, visual sensors were employed, since they are able to provide visual feedback for localization, scene reconstruction and environment inspection. Therefore a low cost web camera and a well known RGB-D sensor were utilized. The cameras were used both as a stereo camera rig and as a monocular visual sensor.

A. Visual Sensors

The PlayStation Eye is a web camera with increased capabilities through specific tweaks. In Figure 3a it is depicted the stereo rig with a canonical configuration, mounted on a 3D printed base, with a baseline of 11 cm, where the camera modules have been disassembled from their casing. This setup has been also utilized as a monocular sensor as stated before. In this article we have considered the left camera as the Monocular sensor of this setup, while the camera resolution used was 640×480 with a frame rate at 30Hz.

Kinect 2, presented in Figure 3b is an RGB-D sensor that includes an RGB camera and it is able to calculate scene depth directly by using a time of flight camera, based on IR transmissions.

B. Experiments Design

The experiments were performed in a tunnel, part of the production area of the iron ore mine in Kiruna. The concept of the experiments was to mount the sensors in a cart and navigate it manually inside the mine tunnel to collect data and validate localization performance, while simulating the aerial platform movements. Each sensor was tested separately and there has not been any fusion among the measurements. This approach has been selected in order to derive the actual capabilities of each sensory system, since it is well known that the corresponding fusion of multiple sensors will always increase the accuracy of the obtained results. In the performed experiments, the lightning source was a portable light projector, which has been fixed on the cart. In Figure 4 the cart carrying the sensory systems is depicted.

For the performed experiments, the visual sensors have been positioned on the cart in the following three configurations:

- Front looking configuration, where the cameras were facing towards the direction of the sensory platform movement. In this case the path was a line followed forward and backward, while covering a distance of 20 meters from the initial starting point inside the production area tunnel.
- Side looking configuration, where the cameras were facing towards the tunnel wall (perpendicular to the movement of the platform) and with a distance of

Fig. 4: Equipment setup during experiments

around 3 meters from it. The translation of the cart was similar to the one described before.

- For both of the previous cases, different heights and angles of the visual sensors on the cart were tested, in order to simulate better the movements of a UAV.

IV. EXPERIMENTAL RESULTS

The post processing of the Kinect recorded data from the field trials, provided a substantial performance with great accuracy in ego-motion estimation and environmental 3D mapping for the Kinect sensor. The visual odometry rate was around 20Hz. The calculated trajectory as well as the obtained corresponding map of the tunnel wall are depicted in the following Figures 5a and 5b.

Regarding the stereo camera system, the localization algorithm provided some satisfactory results but not as accurate compared to Kinect. It should be also noted that after some time the algorithm drifted, affecting majorly the pose estimation of the camera and thus leading to inaccurate measurements with a deviation from the real values. The most successful tests were performed when the cameras were positioned with a direction towards the wall. In this case, the obtained results highlighted that the texture of the environment was adequate for the algorithm to extract features for the ego-motion estimation and also that the cameras were able to calculate the depth information accurately. The tests that were performed with the front look cameras did not succeed in producing long-term results due to the distance from the identified scenes or the lack of texture in the environment. In the presented case, the visual odometry rate was around 20Hz, while the following Figures 6a and 6b present the results from different experimental tests with the stereo cameras.

In general, the obtained results from the stereo camera indicate that the quality of the pose estimation and the 3D map is not perfect. Furthermore, Figure 6 depicting the trajectory and the resulting map by utilizing the stereo camera shows that in some cases the calculations can produce uncertainties in the pose estimation. Nevertheless, it is very promising

that under these challenging conditions and intense darkness, this low cost camera configuration was able to receive and process data. It should be also highlighted that this approach computes odometry only from visual data. Other compact and well optimized stereo systems as well as more elaborate SLAM algorithms could improve the overall performance of the visual sensors.

For the case of the monocular camera, the ORB-SLAM approach was able to provide localization measurements under specific environmental conditions. Some of the test succeeded to localize the camera, but the majority of them failed. Furthermore, in the successful test trials the camera was facing the wall of the tunnel and the cart followed a straight line path. During the initialization process for the camera's position, the algorithm searches for distinctive features in the image. The problem of this approach was that in most of the experiments it was not able to initialize the position of the camera, while it was searching for interesting features, majorly due to poor lightning conditions, low texture environments and abrupt movements inside the tunnel environment. Additionally, the system could not initialize depending on the relative movement of the camera to a planar scene. For the monocular camera, the results of the localization experiments depicted in Figures 7a, 7b clearly shows that the utilized algorithm was able to provide measurements, when it succeeded to initialize. Moreover, it should be noted that the map created by the monocular localization approach was sparse and the camera path followed a straight line. In the presented results, the black points shown in Figures are the past map points, while the red ones are the current map points.

From all the obtained experimental results, a general finding that should be highlighted is the fact that the aforementioned visual techniques require substantial amount of features in the image and texture in the surrounding environment. The most successful experimental results were those concerning the sensors facing towards the tunnel walls, showing clearly that the utilized algorithms work in well textured environments. Unfortunately, it was not possible to acquire ground truth data from the area, since there was no availability of an indoor accurate localization system within the mine. Without a solid base that the ground truth data could provide it is not possible to present a detailed comparison of these methods. Additionally, all the presented results accumulate errors from the actual trajectory due to drift in position and heading.

A video that demonstrates an overview of the obtained results can be reached in [21]

V. FURTHER DEVELOPMENTS

From the previous analysis it is obvious that specific needs for further developments exist, which can be summarized in the following analysis.

A. Flying time limitations

A major practical difficulty for Micro Aerial Vehicles is their limited battery life. In other words, their operating lifes-
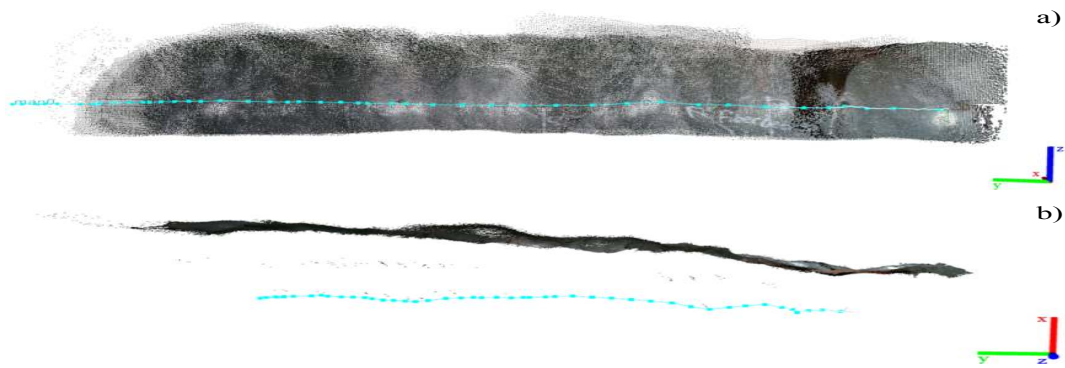
Fig. 5: Kinect sensor trajectory followed and 3D reconstruction. a) Front looking view. b) Top Down view.
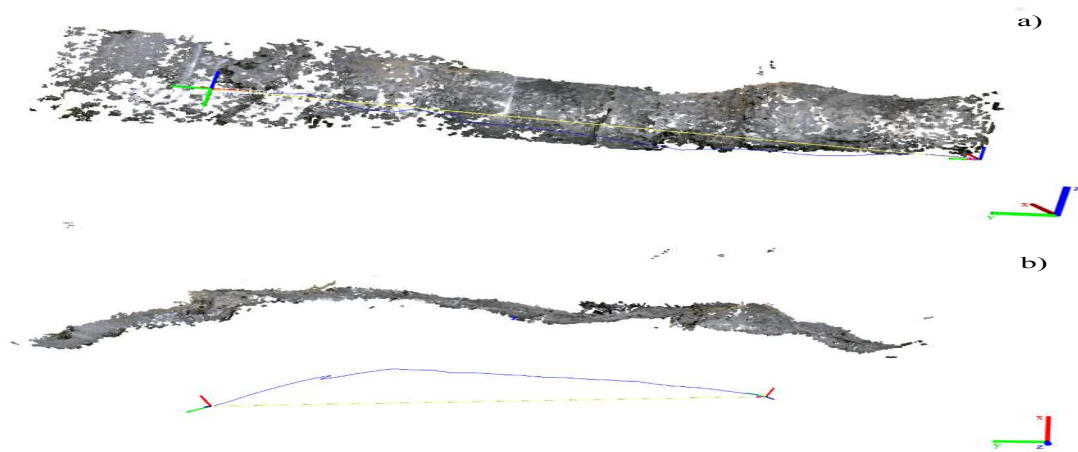

Fig. 6: Stereo sensor trajectory followed and 3D reconstruction. a) Front looking view. b) Top Down view.
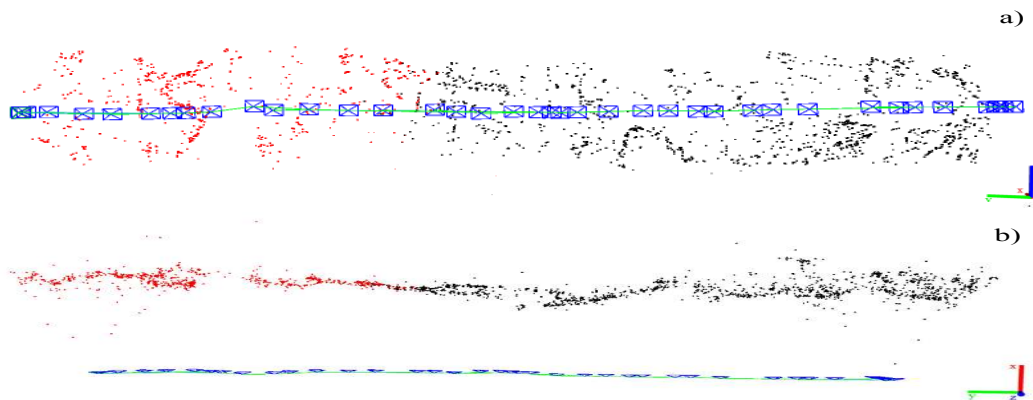

Fig. 7: Monocular sensor trajectory followed and 3D reconstruction. a) Front looking view. b) Top Down view.

pan is restricted and thus is forbidding the proper completion of complicated missions that request long-term exploration. Therefore, it is critical to test thoroughly the battery life in order to overcome their limited amount of time operation and maximize their efficiency. From another point of view, bigger batteries will demand an increased payload, which will result in a bigger and more expensive UAV solution and thus a trade of between flying times and size should be investigated.

B. Light Sources

Visual sensors require a substantial amount of light in the environment in order to provide accurate results. Thereafter, specific illumination strategies should be followed to guarantee proper ambient light without affecting the way cameras perceive the environment. The further developments in this area include the evaluation of different light sources, the corresponding power consumption, the effect on the payload as well as the evaluation of improvement in the performance

of the vision algorithms under variations in illumination.

C. Payload for the UAV

Micro Aerial Vehicles have limitations regarding their payload capabilities. Consequently, it is important to select the visual sensory system depending on the application and to develop the appropriate aerial platform to integrate them. Thus a specific investigation on an optimal and application specific design for the UAV is fundamental as a future development.

D. Tethered autonomous navigation

For solving the battery life problems and the corresponding effects on the payload, an alternative approach to overcome limited battery life is the tethered UAV navigation. Tethered navigation requires the development of specific control schemes for the cable management that could augment the operation time. Such a solution will increase the life span of the battery, however it will also introduce limitations on flying and thus specific control schemes should be developed to overcome the additional limitations.

E. Magnetic dust

Magnetic dust is common to find in iron ore mine areas and pose serious problems. It can also be produced by the thrust of the propellers. Dust blocks majorly the visual information that cameras perceive and therefore deteriorates their performance under such conditions causing inaccurate results. Additionally magnetic dust can damage the electronic hardware and therefore specific solutions to shield and protect the hardware parts are essential for long term operations.

VI. CONCLUSIONS

The aim of this article was to evaluate the performance of current technologies towards visual localization systems in underground mining. The preliminary results from the experiments showed that the stereo cameras could be used for autonomous UAVs but the localization algorithm should not rely solely on them since the results accumulated drift errors. The monocular camera approach showed that it can be used successfully for localization purposes under some constraints and that the results are strongly related to the application that will be employed. Kinect for Windows, provided by far the best results using the same algorithm. The drawback of Kinect is the fact that it is a power consuming sensor with inappropriate size for UAVs that have limited dedicated space for sensors. The advantage of stereo/monocular cameras against Kinect is the fact that consist lightweight sensors that are not power demanding. It should be stressed that within this study the main focus was upon low cost solutions that could be affordable for the end user and potentially increase their reliability in difficult environments like mines. The performance of all the sensors is highly dependent on the surroundings conditions. A major comment that should be made is about sensor fusion concept that can improve the reliability and accuracy of localization algorithms. It is more and more common to merge measurements from IMUs, laser scanners with visual sensors in order to correct pose estimation. The ideal case in autonomous UAVs is to fuse measurements from lightweight sensors in order to get accurate results and guarantee flight time above a specific threshold. In this case more sensors are used to assure the good performance for position and orientation estimation towards autonomous Micro Aerial Vehicle control.

REFERENCES

[1] "Phoenix Aerial Systems TerraHawk T-16 Fixed-wing." [Online]. Available: http://www.phoenix-aerial.com/products/aerial-lidar-systems/tempest-t16/
[2] "AscTec Firefly." [Online]. Available: http://wiki.asctec.de/display/AR/AscTec+Firefly
[3] E. Fresk and G. Nikolakopoulos, "Full quaternion based attitude control for a quadrotor," in *European Control Conference (ECC)*. IEEE, 2013, pp. 3864–3869.
[4] YAMAHA, *RMAX*, http://rmax.yamaha-motor.com.au/features/.
[5] K. P. Valavanis, *Advances in unmanned aerial vehicles: state of the art and the road to autonomy*. Springer Science & Business Media, 2008, vol. 33.
[6] K. Alexis, G. Nikolakopoulos, and A. Tzes, "Switching model predictive attitude control for a quadrotor helicopter subject to atmospheric disturbances," *Control Engineering Practice*, vol. 19, no. 10, pp. 1195–1207, 2011.
[7] G. Nikolakopoulos, T. Gustafsson, P. Martinsson, and U. Andersson, "A vision of zero entry production areas in mines," *IFAC-PapersOnLine*, vol. 48, no. 17, pp. 66–68, 2015.
[8] S.-J. Lee and J.-H. Kim, "Development of a quadrocoptor robot with vision and ultrasonic sensors for distance sensing and mapping," in *Robot Intelligence Technology and Applications 2012*. Springer, 2013, pp. 477–484.
[9] J. Engel, J. Sturm, and D. Cremers, "Scale-aware navigation of a low-cost quadrocopter with a monocular camera," *Robotics and Autonomous Systems*, vol. 62, no. 11, pp. 1646–1656, 2014.
[10] K. Schmid, P. Lutz, T. Tomić, E. Mair, and H. Hirschmüller, "Autonomous vision-based micro air vehicle for indoor and outdoor navigation," *Journal of Field Robotics*, vol. 31, no. 4, pp. 537–570, 2014.
[11] P. Gohl, M. Burri, S. Omari, J. Rehder, J. Nikolic, M. Achtelik, and R. Siegwart, "Towards autonomous mine inspection," in *3rd International Conference on Applied Robotics for the Power Industry (CARPI)*. IEEE, 2014, pp. 1–6.
[12] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (slam): Part ii," *IEEE Robotics & Automation Magazine*, vol. 13, no. 3, pp. 108–117, 2006.
[13] D. Demirdjian and R. Horaud, "Motion–egomotion discrimination and motion segmentation from image-pair streams," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 53–68, 2000.
[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
[15] M. Labbe and F. Michaud, "Appearance-based loop closure detection for online large-scale and long-term operation," *IEEE Transactions on Robotics*, vol. 29, no. 3, pp. 734–745, 2013.
[16] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[17] R. Mur-Artal, J. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," *arXiv preprint arXiv:1502.00956*, 2015.
[18] G. Klein and D. Murray, "Parallel tracking and mapping for small ar workspaces," in *6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2007, pp. 225–234.
[19] D. Filliat, "A visual bag of words method for interactive qualitative localization and mapping," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2007, pp. 3921–3926.
[20] "Robot Operating System." [Online]. Available: http://www.ros.org/
[21] "Video presenting the experimental results of this study." [Online]. Available: https://www.youtube.com/watch?v=g9KGKLo2qJ0