Photogrammabot: An Autonomous ROS-Based Mobile Photography Robot for Precise 3D Reconstruction and Mapping of Large Indoor Spaces for Mixed Reality

Soroosh Mortezapoor*
TU Wien, Vienna, Austria

Christian Schönauer†
TU Wien, Vienna, Austria

Julien Rüggeberg‡
Illusion Walk KG, Berlin, Germany

Hannes Kaufmann§
TU Wien, Vienna, Austria

ABSTRACT

Precise 3D reconstruction of environments and real objects for Mixed-Reality applications can be burdensome. Photogrammetry can help to create accurate representations of actual objects in the virtual world using a high number of photos of a subject or an environment. Photogrammabot is an affordable mobile robot that facilitates photogrammetry and 3D reconstruction by autonomously and systematically capturing images. It explores an unknown indoor environment and uses map-based localization and navigation to maintain camera direction at different shooting points. Photogrammabot employs a Raspberry Pi 4B and Robot Operating System (ROS) to control the exploration and capturing processes. The photos are taken using a point-and-shoot camera mounted on a 2-DOF micro turret to enable photography from different angles and compensate for possible robot orientation errors to ensure parallel photos. Photogrammabot has been designed as a general solution to facilitate precise 3D reconstruction of unknown environments. In addition we developed tools to integrate it with and extend the Immersive Deck™ MR system [23], where it aids the setup of the system in new locations.

Keywords: Photogrammetry, 3D reconstruction, automation, autonomous mobile robot, ROS, mixed-reality, fiducial marker tracking.

Index Terms: Computer systems organization—Embedded and cyberphysical systems—Robotics—Robotic autonomy; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality; Computing methodologies—Artificial intelligence—Computer vision—Reconstruction;

1 INTRODUCTION

The term Mixed Reality (MR) describes "a particular subclass of VR related technologies that involve the merging of real and virtual worlds" [21] including consistent visual presentation to a user through an appropriate display device. At the same time, there can be interactions between the virtually presented objects, the actual objects, and the user. One of the main technical challenges in Mixed Reality (MR) is for a system to know the geometry of the physical world relative to the position of the display system [8]. In addition to that, creating high-quality content for MR, which visually blends well with the actual environment, is burdensome and requires considerable effort. Photogrammetry [20] can aid these goals by precisely reconstructing textured 3D models of environments and objects [27]. The proposed robotic system, Photogrammabot, which is explained thoroughly later here and is shown in Fig. 1 can facilitate photogrammetry. Moreover, photogrammetry can help the calibration of tracking systems by mapping positions of fiducial markers as used in the Immersive Deck™ [23], shown in Fig. 2, and briefly explained later in the following sections, natural features, sensor systems or determine boundaries of the physical space.

Figure 1: Photogrammabot in action capturing photos of an indoor environment for 3D reconstruction using photogrammetry.

However, photogrammetry is time-consuming and requires hundreds or even thousands of photos and expertise regarding images' perspectives and alignment relative to the targeted physical environment. The number of correct photos correlates with the quality of the result [6], thus directly impacting the precision of the reconstructed environment. Above mentioned points imply that the photography stage of photogrammetry is of high importance to obtain high-quality 3D reconstructed models, while in nature, this step consists of repetitious tasks, relocating and shooting photo. Robots are proven to be significantly helpful for such repetitious jobs, and they are employed for similar tasks such as monitoring and surveillance successfully.

Autonomous mobile robots nowadays are utilized in a wide variety of tasks in both indoor and outdoor environments. They can operate in conditions that are not suitable for humans and can perform repeating tasks accurately. Moreover, toy robots are becoming affordable, consequently more popular than ever to use in research or industrial projects such as monitoring, mapping, camerawork, etc. Many robots belonging to diverse categories of robots, whether commercially used or built and utilized in research, nowadays enjoy the benefits of Robot Operating System (ROS) [26]. ROS empowers robotic researchers and developers by offering a powerful set of

*e-mail: soroosh.mortezapoor@tuwien.ac.at
†e-mail: christian.schoenauer@tuwien.ac.at
‡e-mail: julien@illusion-walk.com
§e-mail: hannes.kaufmann@tuwien.ac.at

Figure 2: Immersive Deck™ [23]. (a) An immersed user localized using the upward-directed camera having a fisheye lens mounted on the HMD. The location of the user is determined using the current view of the camera of the fiducial markers on the ceiling compared to the marker map obtained from the 3D reconstructed model. (b) Multiple immersed users in group walkable-VR while being localized using Immersive Deck™ system.

ready-to-use libraries and tools to control diverse types of robots while being able to run on an IoT computer such as Raspberry Pi (RPi) [25].

In this work, we propose a ROS-based robotic setup called Photogrammabot to facilitate precise 3D reconstruction of an unknown indoor environment. Photogrammabot, which is presented in Fig. 1, can take a very high number of aligned photos without any human effort over hours of operation. In the following sections, the hardware and software specifications of the system are described. Subsequently, we demonstrate the results obtained from the evaluation of the system, discuss the significance of the results compared to manually taken photos by a human, and explain future work.

2 Related work

Mobile Robots are gaining popularity in diverse domains to perform various tasks. Different solutions for the 3D reconstruction of real-world objects and environments using robots are proposed depending on the application. The majority of the robotic systems employed for such tasks are manually controlled, commanded and guided by humans. Toschi et al. in [32] propose such a remotely-controlled so-called sensorized car for remote mapping of an environment.

Autonomous Mobile Robots, AMRs, in contrast, are those which are capable of navigating and exploring an environment, and in this domain, acquiring required data for 3D reconstruction without any human intervention. Despite being on their own coordinating their tasks, such systems can still enjoy the benefits of some prior information about the environment, such as a pre-determined flight path for a UAV in a monitoring or inspection task. In the case of photography or scanning for 3D reconstruction, scan planning and execution need to be considered and tackled during the navigation phase by the help of prior knowledge to pick the photography poses. On the other hand, the challenge of finding the best viewpoint for different purposes, including 3D reconstruction, has been widely studied. As an alternative to predefined photography poses, the best viewpoints can be calculated using the solutions of the Next Best View (NBV) problem systematically [3, 9, 18, 19, 22] at the cost of more essential computations, more sensors, and as a result higher power consumption.

The majority of the automated 3D reconstruction systems that utilize robots gather information in the form of colored point cloud data using 3D laser scanners or RGBD cameras such as those presented in [4, 5, 7, 13, 31]. Adan et al. in [2] propose an integrated system that provides detailed as-is semantic 3D models of buildings using a five level semantic model, whose first level is created using an autonomous robotic scanning platform that collects data regarding the scene and generates a point cloud that is later structured in a semantic point cloud model containing indoor, clutter and outlier point clouds. One of the most visible problems with reconstructing an environment using point clouds is that the final textures need to be created out of the point cloud data, which results in disparity between the actual environment and the 3D reconstructed version, such that the reconstructed model looks rather like a collection of artificial game objects, as it can be seen in the results of [2].

Potthast and Sukhatme in [24] propose a 3D reconstruction system using a PR2 robot with a range of high-end sensors such as tilt laser scanner, stereo cameras, RGBD sensors, inertial sensors, etc. It follows the same principle of gathering information in the form of 3D point cloud data. Compared to PR2, our proposed robotic system is more affordable thanks to its simplicity and fewer sensors employed. It is also significantly smaller and lighter compared to PR2, which makes it easier to carry to the new scanning sites as needed by the system explained in [23].

Additionally, many people proposed robot-aided data acquisition for 3D reconstruction using Unmanned Aerial Vehicles, UAVs. Such solutions can be found in [1, 11, 16, 17]. To construct a 3D model of an environment, many appropriate images that cover the environment entirely must be acquired. The most important limiting factor when it comes to using UAVs for 3D reconstruction data acquisition is the maximum possible operation time. Due to load constraints, UAVs cannot carry heavy batteries, and therefore their flying time is very limited compared to the total operation time of a ground robot. Another crucial factor for the best possible results of 3D reconstruction using photogrammetry is the quality of the obtained photos. The proposed ground robot in this work can handle different cameras including cameras with Full-frame or even Medium format sensors, which can increase the quality of the photos considerably, especially in indoor environments where there might be not enough light for normal camera sensors to take sharp noiseless pictures. Needless to say that such cameras with large sensors need heavy and relatively large objectives, and are usually too heavy and power-consuming to be carried and operated on a drone. It should be added that a drone that can carry a camera needs powerful motors, which can cause heavy wind in indoor environments, which can consequently cause issues.

3 System Design and Implementation

Photogrammabot has been designed as a general solution to facilitate precise 3D reconstruction of unknown environments. In addition we developed tools to integrate it with and extend the Immersive

Deck™ MR solution [23], where it also aids setup of the system in new locations. Photogrammabot comprises several robotic components to achieve the highest possible autonomy, system stability, and reliability while keeping the final product affordable. In this section, the hardware and software specifications of the system are explained.

3.1 Hardware

Photogrammabot employs an RPi 4B as the processing unit. In the following, the essential hardware components of Photogrammabot are listed. Fig. 3 illustrates the internal arrangement of the elements. Although the components utilized in this setup and listed below are well-known among robotic researchers and developers and widely used in similar robotic configurations, they can be replaced with similar ROS-compatible components if necessary.

- **Kobuki base [34] with 4400mAh battery**, used as both the mobile base and the power source for the other components. Kobuki provides a USB interface for two-way communication to the onboard computer such that the computer can read specific types of data such as battery and bumper status, current velocity, and odometry data and control the velocity and acceleration of the motors, consequently the robot. Kobuki drivers and libraries for ROS expose these data to ROS and its working nodes and offer standard interfaces to control the movement of the motors, which could be imported and used by other high-level ROS nodes and algorithms to control the locomotion.

- **RPi 4B 8GB RAM [10]** and touchscreen display kit used as the sole onboard computer. RPi controls the whole operation by reading sensor data and producing proper actuation commands.

- **PhantomX micro turret** with ArbotiX-M Robot controller [28] to enable reorienting the camera in-place and photography from different angles without having to relocate the entire robot using its mobile base.

- **Sony DSC-RX100M4 Camera** for photography of the environment.

- **Sony Camera Multiport adapter with charging capability**, used to control the focus and shutter release functionalities of the Sony camera from the onboard computer and provide power to the camera externally.

- **Rplidar A3M1 25m range laser range scanner (LiDAR)** which is an affordable 360-degree LiDAR utilized for mapping, localization, navigation, and collision avoidance purposes. [30]

3.2 Software

We utilized ROS Melodic to accelerate the development process due to its richness of the libraries and the possibility of operating on IoT computers such as RPi. Like all other ROS distributions, ROS Melodic depends on a specific version of Ubuntu, here version 18.04. However, only Ubuntu 18.04 Core is available for RPi, which comes with no UI. Alternatively, Ubuntu-mate provides an almost similar environment to Ubuntu but can be installed and run on an RPi. Therefore, Ubuntu-mate 18.04 is installed as the operating system. An installation of ROS Melodic makes it possible to benefit from a wide variety of ROS-powered algorithms and plugins, called ROS packages, such as ready-to-use localization and navigation solutions, ROS controllers, and hardware abstraction tools. ROS makes it possible to make use of its libraries, and further expand the functionalities using Python or C++ code. The complete source code of Photogrammabot, as well as the hardware and setup instructions are publicly accessible on Github[1].

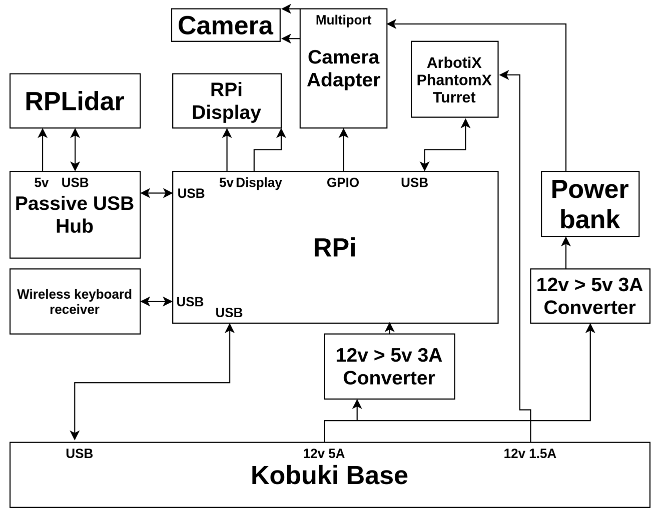

Figure 3: Hardware component diagram of Photogrammabot.

3.2.1 Hardware abstraction and control

General Purpose Input/ Output (GPIO) of the onboard RPi triggers the focus function and shutter release, respectively. A wide range of digital cameras enhanced with an interface for wired external shutter release controllers offer standard 3-pin interfaces consisting of ground, focus, and shutter release. Nevertheless, the camera used in this project needs a Sony Multiport adapter that we built based on the pin-out information provided by Sony. The adapter can provide external power drawn from the added power bank and control the photo shooting function simultaneously.

3.2.2 Localization and Mapping

As mentioned in the previous sections, the robot can explore an unknown indoor environment using a greedy frontier-based exploration method [15] and navigate through it to maintain flexibility. Nevertheless, a map is necessary to plan for the photography points in the environment. Furthermore, the robot must localize itself and know its precise position to follow the produced plan. For this purpose, the famous *gmapping* ROS package [14] [29] is used. *gmapping* is capable of producing a 2D map based on the LiDAR data and localizing the robot simultaneously.

3.2.3 Navigation and Exploration

We used *move_base* ROS package for navigation purposes. The *Eband* local planner that can be plugged into the *move_base* package helps the robot avoid collisions to both static and dynamic obstacles and turned out to be a fit for this purpose due to its simplicity and lightness from the processing power consumption aspect.

explore_lite, a greedy frontier-based exploration ROS package, is used to explore the unknown indoor environment. *explore_lite* enables Photogrammabot to explore the environment autonomously. It continuously explores the environment until no more frontiers with potential for further exploration exist on the map.

3.2.4 Camera angle control

To control the camera's pan and tilt angles, we enhanced the robot with a PhantomX micro turret having an ArbotiX-M Robotcontroller. The ROS-enabled ArbotiX-M Robotcontroller can change the pan

[1] `https://github.com/smortezapoor/photogrammabot_pub`

and tilt angles of the micro turret, both ranging from $-150°$ to $+150°$.

3.2.5 Orchestration

A ROS package is developed to orchestrate the operations and automate the entire process. The subtasks are orchestrated in the following order:

1. Explore and create a 2D map
2. Create a plan for shooting points.
3. Go to the first shooting point in the path
4. Follow the path. For each shooting point[2]:
   (a) Stop at the point
   (b) For each predefined configurable shooting pan and tilt:
      i. Take a photo

In Fig. 4 a simulated Photogrammabot in the middle of the photography operation is presented. Fig. 4.*b* shows a simulated room in Gazebo [12] created from reality with some additional obstacles randomly distributed in the scene. As the Photogrammabot's structure is defined using a description file which is recognizable by Gazebo as well, Gazebo can accurately simulate the behavior of the sensors, such as the utilized LiDAR and the actuators of the robot. Fig. 4.*b* depicts the processed and visualized data of the simulated robot's sensors in *Rviz*. As shown in this figure, a map is generated after exploring the environment. The planned photography points are illustrated on the map using black numbered circles with an arrow that specifies the robot's expected orientation at the point. The photography points with green arrows pointing outward from them are already traversed by the robot and each green arrow indicates the direction of a successfully captured photograph at the point. A yellow circle represents the robot almost in the middle of the environment while following its next goal.

The plan created in step (2) of the above-mentioned orchestration algorithm determines the photography points in the already explored environment based on the generated map, meaning the positions at which the robot should capture photos. A square wave-shaped path is generated using the map data to optimize the travel length of the robot to cover the whole area and consequently reduce the power consumption. The distance of each photography point to its preceding one is $1m$ unless the map shows that the desired photography point is occupied i.e. there is a column in that location in the room.

The configuration of the path planning algorithms of Photogrammabot guarantees the orientation of the robot at each photography point to be within $-10°$ to $+10°$ tolerance range compared to the heading of the robot when the operation was first started. Although the orientation of the robot can vary up to $20°$ at each photography point compared to its previous one, thanks to the localization mechanism, the robot knows its precise pose. It uses this discrepancy in the orientation of the robot's heading as a compensation value for the micro turret pan angle to ensure parallel sets of photos, which is required for photogrammetry.

In the subtask 4.*b* of the orchestration algorithm, a configuration file allows specifying the number of photos at each photography point and their horizontal and vertical angles, respectively.

[2] As the robot is capable of handling dynamic obstacles, reaching to each photography point has a 60 seconds timeout and if the robot cannot reach the destination point before this timeout is expired i.e. due to a dynamic obstacle moving to the destination point after planning, the photography point is abandoned and the next point along the path is followed.

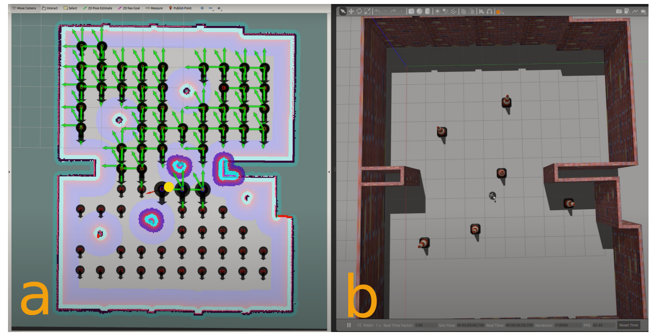

Figure 4: A simulated model of Photogrammabot in the middle of the photography operation. (a) The visualized data of the Photogrammabot in ROS visualization tool, *Rviz* as a 2D map. The yellow circle in the middle represents the robot. (b) The simulated robot and the indoor environment with some static obstacles in the environment in Gazebo.

3.2.6 Photogrammetry and 3D reconstruction

For processing the photos and 3D reconstruction of the environment, RealityCapture [6] is employed. RealityCapture is a well-known software for photogrammetry, capable of generating 3D models from many images. The generated 3D models can then be used in Virtual Reality and Mixed-Reality applications. One crucial precondition for a successful reconstruction is that the photos have sufficient overlap to be registered by matching features between images. At the same time, the images have to be taken from close enough proximity for details to be reconstructed with sufficient detail. A human photographer needs experience and tools to select suitable positions and perspectives for capturing images. For our automated approach, preferable spacing and angles between photos can be easily determined mathematically from a targeted overlap percentage and configured in the system as stated before. The robot will then follow that configuration with high accuracy, taking shake-free pictures with good repeatability.

3.2.7 Immersive Deck™ MR system

Immersive Deck™ is an immersive MR system first introduced by Podkosova et al. in [23]. It allows several users to simultaneously explore large Virtual Environments (VE) by natural walking and interact with each other and real or virtual objects in an intuitive way, as demonstrated in Fig. 2 (b). Immersive Deck™ is largely built from low cost off-the-shelf hardware, while the workflow and algorithms for large-scale marker-based tracking makes it easy to set up. Photogrammabot and our customized workflow facilitate the setup process by autonomously mapping the real environment and fiducial marker positions. For that purpose we implemented a tool that automatically detects the 2D marker positions in the Photogrammabot photos and stores them along with their ids in a .csv file for further processing. In Reality Capture the user loads the data and starts the reconstruction. Reality Capture along with the 3D reconstructed textured mesh automatically computes the 3D positions of the markers, which the user in turn can export and feed into our post-processing tool, which generates a marker map readable by the Immersive Deck™ marker tracking component. Utilizing this marker map the Immersive Deck™ coordinate system can be precisely registered with the real world, while the reconstructed 3D models of the environment can be imported to be used in MR experiences.

4 Experimental Results

We evaluated the robot and the reconstruction pipeline in a room of over 200 square meters, where the robot captured 1386 photos within four hours. The test environment contained 400 square fiducial markers with 21.6*cm* side length evenly distributed on the ceiling and upper parts of the walls. The marker coordinates were precisely located by manually post-processing laser scanner data with a target accuracy of $2-5mm$ for the Triple-Track™ tracking system [33], extending upon the marker-based method described in [23]. The compared corner positions show only a tiny deviation after aligning the 3D reconstructed model with the laser-scanned 3D marker positions by transformation into a coordinate system defined by three markers. Our automated post-processing step removes outliers, easily identified by wrong marker dimensions. The deviation between 3D reconstructed corners and the laser-scanned "ground-truth" averages 2.2*cm* (SD 3.3*cm*). Only two percent of the markers were removed as outliers, indicating a good coverage of the targeted area. The length of the room as measured from the 3D model deviates by just 1.3*cm* from the measurement of the laser scanner, which is slightly over 0.05% of the 25.44*m* room length. On this scale, the error is negligible for most MR applications. Furthermore, the resulting subjective tracking experience of the experimenters is just as good as with the laser-scanned marker-map. Finally, the reconstructed mesh and texture appear detailed and realistic from visual inspection and can be integrated into an MR experience. Fig. 5 (b) demonstrates the 3D reconstructed model of the actual room after photogrammetry, while Fig. 5 (a) shows the same part of the room in a normal camera photo taken by the Photogrammabot.

It is noteworthy that to reduce the power consumption of the mechanical parts of the robot, meaning the wheels and the servos of the PhantomX micro turret, and lower the risk of introducing errors, the working velocity and the maximum forward and backward acceleration of the robot and the micro turret are limited to 20% of their maximums. These values can be increased for operation in larger environments using configuration files, resulting in significantly shorter required time-to-complete, at the cost of higher battery consumption rate, and more required processing power by the SLAM algorithm to cancel the localization error, specially when the area has a slippery floor that can cause the robot to slip due to high translational and rotational accelerations. The other reason for setting the limits to 20% is that our use cases for the Photogrammabot are not time-sensitive. The priority was extended battery running time, lower robot maintenance costs and effort, more accurate images, and better photogrammetry results. The current settings allow the Photogrammabot to run around 6 hours.

Photogrammabot can save several hours of tedious, repetitious human labor per reconstruction. For a human photographer with the same camera as the robot's, the average time needed for each photo considering the relocation to cover the entire area is measured around 10 seconds, resulting in a total operation duration similar to the required time by Photogrammabot for the same place at 20% velocity and acceleration limit. Additionally, specific measuring tools such as a compass should be utilized by him/her for taking aligned photos, which increases the time per photo even more. In cases where insufficient lighting or precision requirements make the use of a tripod or other tools necessary, required time can increase even further. Although a human's needed amount of time to accomplish the operation can be reduced to some extent after training and practice, these still need investing time and human effort. In addition, for a better overlap of ceiling images without distortion from a wide-angle lens, the photos are better to be taken from as close to the ground as possible. The camera's height in Photogrammabot is below 40*cm* from the ground, while an average camera holding height for a human is considerably higher. Needless to mention, holding the camera near the ground can significantly increase the time and inconvenience for the human photographer. Photogrammabot, on the other side, requires under one minute of manual intervention to set up and start, while it automatically produces high-quality photos with predictable results of the photogrammetric reconstruction, thus

Figure 5: (a) A sample photo, taken by the Photogrammabot from the room for photogrammetry. (b) The rendered result of photogrammetry of the same room.

saving many hours of human work time per scan in many situations. Furthermore, it is far more energy-efficient for remote sites to ship a comparatively small robot that needs no special training for starting the operation than have human photographers travel by plane to the target site.

5 Conclusion and Future Work

We have presented Photogrammabot, an autonomous robotic setup to facilitate precise 3D reconstruction of an unknown indoor environment by capturing a very high number of aligned photos from predefined angles. The evaluation results presented in the previous section demonstrated the fitness of Photogrammabot for the mentioned purpose from both providing good quality input images for photogrammetry to build high-quality 3D reconstructed content for MR, and reducing human effort to collect them. In addition, Photogrammabot has been shown to be suitable for aiding mapping of fiducial markers for tracking in the Immersive Deck™ MR system.

Although the time spent by the robot for the photography operation stayed roughly similar to a human photographer in our experiment, the human effort spent for the process was decreased by a factor of two orders of magnitude. As mentioned before, increasing the velocity and the acceleration limits of the robot and its integrated micro turret can make the process multiple times faster for the scenarios where shorter time-to-complete is desirable. Photogrammabot can save several hours of time and effort for MR developers and researchers and ensures the quality of the final reconstructed space supported by its precision of image capturing orientation. It is affordable, easy to build and assemble using the given instructions and the publicly shared code repository, and effective not only for one-room environments, but an entire connected floor thanks to its frontier-based exploration algorithm and its ability for collision-free navigation.

In future work we want to conduct comparative studies to determine the final quality of the 3D reconstructed model and the ease of use of Photogrammabot compared to the manually-capturing-photo method and current 3D scanning with LiDAR. Furthermore, we will perform tests in multi-room environments and adapt the algorithms where necessary in future work. Ultimately, a complement to this can be adding the capability of going up or down the stairs for a multi-story 3D reconstruction by augmenting the Kobuki base, or replacing it with a mobile base with stair climbing capabilities. For larger environments, the battery capacity can be increased, as Kobuki base is claimed to be able to handle 5Kg payload while in this setup the weight of the extra components used to build the Photogrammabot is under 2Kg. As a result, the rest 3Kg unused payload capacity can be dedicated to adding extra battery to prolong the operation time, or supply power to new components.

Finally, an object scanning mode could be developed, focusing the robot movement and image capture on one physical object or a small collection of objects to be reconstructed as content in MR experiences.

6 ACKNOWLEDGMENTS

This work has been partially funded by the European Union's Horizon 2020 research and innovation programme within the BRIDGES project under grant agreement No 952043. Moreover, we gratefully acknowledge the support by grant F77 (SFB "Advanced Computational Design", Sub-project 5) of the Austrian Science Fund FWF.

REFERENCES

[1] A. F. Abate, L. De Maio, R. Distasi, and F. Narducci. Remote 3D face reconstruction by means of autonomous unmanned aerial vehicles. *Pattern Recognition Letters*, 147:48–54, July 2021. doi: 10.1016/j.patrec.2021.04.006

[2] A. Adán, B. Quintana, S. A. Prieto, and F. Bosché. An autonomous robotic platform for automatic extraction of detailed semantic models of buildings. *Automation in Construction*, 109:102963, Jan. 2020. doi: 10.1016/j.autcon.2019.102963

[3] P. S. Blaer and P. K. Allen. Two Stage View Planning for Large-Scale Site Modeling. In *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, pp. 814–821, June 2006. doi: 10.1109/3DPVT.2006.143

[4] P. S. Blaer and P. K. Allen. Data acquisition and view planning for 3-D modeling tasks. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 417–422, Oct. 2007. ISSN: 2153-0866. doi: 10.1109/IROS.2007.4399581

[5] D. Borrmann, A. Nüchter, M. akulović, I. Maurović, I. Petrović, D. Osmanković, and J. Velagić. A mobile robot based system for fully automated thermal 3D mapping. *Advanced Engineering Informatics*, 28(4):425–440, Oct. 2014. doi: 10.1016/j.aei.2014.06.002

[6] CapturingReality. Realitycapture mapping and 3d modeling photogrammetry software. `https://www.capturingreality.com/`, 2021. Accessed: 2021-11-12.

[7] B. Charrow, G. Kahn, S. Patil, S. Liu, K. Goldberg, P. Abbeel, N. Michael, and V. Kumar. Information-Theoretic Planning with Trajectory Optimization for Dense 3D Mapping. In *Robotics: Science and Systems XI*. Robotics: Science and Systems Foundation, July 2015. doi: 10.15607/RSS.2015.XI.003

[8] E. Costanza, A. Kunz, and M. Fjeld. *Mixed Reality: A Survey*, pp. 47–68. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. doi: 10.1007/978-3-642-00437-7_3

[9] E. Dunn, J. van den Berg, and J.-M. Frahm. Developing visual sensing strategies through next best view planning. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4001–4008, Oct. 2009. ISSN: 2153-0866. doi: 10.1109/IROS.2009.5354179

[10] R. P. Foundation. Raspberry pi 4b. `https://www.raspberrypi.org/products/raspberry-pi-4-model-b/"`. Accessed: 01.08.2021.

[11] K. Fudala and R. J. Bialik. The use of drone-based aerial photogrammetry in population monitoring of Southern Giant Petrels in ASMA 1, King George Island, maritime Antarctica. *Global Ecology and Conservation*, 33:e01990, Jan. 2022. doi: 10.1016/j.gecco.2021.e01990

[12] Gazebo. Gazebo, an open-source 3d robotics simulator. `http://gazebosim.org/`, 2021. Accessed: 2021-11-12.

[13] R. Grabowski, P. Khosla, and H. Choset. Autonomous exploration via regions of interest. In *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*, vol. 2, pp. 1691–1696 vol.2, Oct. 2003. doi: 10.1109/IROS.2003.1248887

[14] G. Grisettiyz, C. Stachniss, and W. Burgard. Improving grid-based slam with rao-blackwellized particle filters by adaptive proposals and selective resampling. In *Proceedings of the 2005 IEEE international conference on robotics and automation*, pp. 2432–2437. IEEE, 2005.

[15] J. Hörner. Map-merging for multi-robot system, 2016.

[16] G. Y. Jeong, T. N. Nguyen, D. K. Tran, and T. B. H. Hoang. Applying unmanned aerial vehicle photogrammetry for measuring dimension of structural elements in traditional timber building. *Measurement*, 153:107386, Mar. 2020. doi: 10.1016/j.measurement.2019.107386

[17] R. Kurazume, S. Oshima, S. Nagakura, Y. Jeong, and Y. Iwashita. Automatic large-scale three dimensional modeling using cooperative multiple robots. *Computer Vision and Image Understanding*, 157:25–42, Apr. 2017. doi: 10.1016/j.cviu.2016.05.008

[18] K.-L. Low and A. Lastra. Efficient Constraint Evaluation Algorithms for Hierarchical Next-Best-View Planning. In *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, pp. 830–837, June 2006. doi: 10.1109/3DPVT.2006.52

[19] J. Maver and R. Bajcsy. Occlusions as a guide for planning the next view. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(5):417–433, May 1993. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence. doi: 10.1109/34.211463

[20] E. M. Mikhail, J. S. Bethel, and J. C. McGlone. Introduction to modern photogrammetry. *New York*, 19, 2001.

[21] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[22] R. Pito. A solution to the next best view problem for automated surface acquisition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):1016–1030, Oct. 1999. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence. doi: 10.1109/34.799908

[23] I. Podkosova, K. Vasylevska, C. Schönauer, E. Vonach, P. Fikar, E. Broneder, and H. Kaufmann. Immersivedeck: A large-scale wireless vr system for multiple users. In *Proceedings of the 9th Workshop on Software Engineering and Architectures for Realtime Interactive Systems SEARIS 2016*. IEEE Xplore Digital Library, 2016.

[24] C. Potthast and G. S. Sukhatme. A probabilistic framework for next best view estimation in a cluttered environment. *Journal of Visual Communication and Image Representation*, 25(1):148–164, Jan. 2014. doi: 10.1016/j.jvcir.2013.07.006

[25] A. Prakash and R. Walambe. Military surveillance robot implementation using robot operating system. In *2018 IEEE Punecon*, pp. 1–5, 2018. doi: 10.1109/PUNECON.2018.8745403

[26] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, A. Y. Ng, et al. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, vol. 3, p. 5. Kobe, Japan, 2009.

[27] F. Remondino. Heritage recording and 3d modeling with photogrammetry and 3d scanning. *Remote sensing*, 3(6):1104–1138, 2011.

[28] T. Robotics. Phantomx micro turret. `https://www.trossenrobotics.com/p/phantomX-robot-turret.aspx`. Accessed: 01.08.2021.

[29] ROS. Gmapping. `http://wiki.ros.org/gmapping`. Accessed: 01.08.2021.

[30] Slamtec. Rplidar a3m1. `https://www.slamtec.com/en/Lidar/A3`. Accessed: 01.08.2021.

[31] H. Surmann, A. Nüchter, and J. Hertzberg. An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments. *Robotics and Autonomous Systems*, 45(3):181–198, Dec. 2003. doi: 10.1016/j.robot.2003.09.004

[32] I. Toschi, P. Rodríguez-Gonzálvez, F. Remondino, S. Minto, S. Orlandini, and A. Fuller. Accuracy evaluation of a mobile mapping system with advanced statistical methods. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-5-W4, pp. 245–253. Copernicus GmbH, Feb. 2015. ISSN:

1682-1750. doi: 10.5194/isprsarchives-XL-5-W4-245-2015
[33] I. Walk. Triple-track system. https://www.illusion-walk.com/technology/. Accessed: 12.11.2021.
[34] YUJINROBOT. Kobuki base robot. http://kobuki.yujinrobot.com/about2/". Accessed: 01.08.2021.