COMBINING LIDAR SLAM AND DEEP LEARNING-BASED PEOPLE DETECTION FOR AUTONOMOUS INDOOR MAPPING IN A CROWDED ENVIRONMENT

D. Tiozzo Fasiolo, E. Maset, L. Scalera, S. O. Macaulay, A. Gasparetto, A. Fusiello*

Polytechnic Department of Engineering and Architecture (DPIA), University of Udine, Udine, Italy
(diego.tiozzo, eleonora.maset, lorenzo.scalera, sadiqolayiwola.macaulay, alessandro.gasparetto, andrea.fusiello)@uniud.it

Commission I, ICWG I/IV

KEY WORDS: Mobile Robotics, Laser Scanning, Simultaneous Localization and Mapping, People Detection, Deep Learning, Indoor Mapping.

ABSTRACT:

In this paper, we present a mapping system based on an autonomous mobile robot equipped with a LiDAR device and a camera, that can deal with the presence of people. Thanks to a deep learning approach, the position of humans is identified and a new surveying path is planned that brings the robot to scan occluded areas, so as to obtain a complete point cloud of the environment. Experimental results are performed with a wheeled mobile robot in different crowded scenarios, showing the applicability of the proposed approach to perform an autonomous survey avoiding occlusions and automatically removing from the map noisy and spurious objects caused by people presence.

1. INTRODUCTION

The contamination between Photogrammetry and Computer Vision in surveying is a process that started at least a decade ago and is now in a mature, though not yet complete, phase. We are also currently witnessing the entry of techniques borrowed from robotics into the field of surveying, as for instance portable Mobile Mapping Systems (MMSs). In fact, the case of a vehicle or a human operator equipped with laser sensors or cameras moving in the environment with the aim of surveying is not at all dissimilar to that of an autonomous vehicle (robot) moving in an unknown environment with the aim of obtaining a spatial representation (mapping). One difference is that the robot, in order to move, must simultaneously solve the problem of localization in that unknown environment, hence the problem known as SLAM (Simultaneous Localisation And Mapping). In fact, portable MMSs coupled with SLAM technology have recently revolutionised modelling in global navigation satellite system (GNSS)-denied environments, driven by the growing demand for up-to-date, high-resolution 3D models of buildings and infrastructures (Maset et al., 2021).

Nowadays, a surveyor can simply carry the mapping device while walking through the area of interest to obtain a point cloud of the surrounding scene, which serves as a fundamental starting point for the creation of digital twins and the as-built Building Information Model (BIM) reconstruction (Rausch and Haas, 2021).

A further step towards the automation of survey operations can be made thanks to robotics and artificial intelligence. Mobile robots, also known as Unmanned Ground Vehicles (UGVs), equipped with mapping systems, indeed, are increasingly applied for the digitization of buildings (Adán et al., 2019, Maset et al., 2022). Steered from remote or autonomously performing navigation, these solutions will represent in the near future a fundamental aid for a more efficient, fast and accurate survey. Examples of applications of mapping systems based on mobile robots are the indoor environment monitoring and control, as in (Park and Hwang, 2020), and the robotic object search by leveraging metric-topological map, as in (Zhang et al., 2021). Other applications include BIM-integrated collaborative robotics for building construction and maintenance (Asadi et al., 2018), functional support for occupancy analysis of building, and autonomous disinfection in challenging situations, such as the COVID-19 pandemic (Giusti et al., 2021).

Although many autonomous scanning systems have been proposed in the literature, several problems still need to be addressed and significant improvements are required. According to (Adán et al., 2019), efforts should focus on implementing computationally more efficient path planning algorithms, providing quantitative evaluations on the accuracy of retrieved 3D models, and improving the degree of autonomy of mobile robotic platforms in complex scenes. Moreover, the developed robotic mapping platforms are mostly tested in indoor scenarios which are usually empty, static, and where there are no people. Indeed, during the scanning of a building or an indoor environment, one of the principal sources of uncertainty and data gaps is occlusion, such as pieces of furniture, shelves, or even human beings, as in the case considered in this work. To avoid occlusions, the operators usually manually scan the environment from many different viewpoints and thus generate a fused point cloud. This could be a serious limitation in the applicability of the system, since in many situations the presence of people during surveying operations cannot be avoided. For this reason, in the context of autonomous mapping systems based on mobile robots, it is convenient to develop techniques to avoid occlusions and remove noisy and spurious objects from the resulting map. As for instance, in (Hähnel et al., 2003), a probabilistic approach is implemented to track multiple people and to incorporate the estimates of the tracking technique into the mapping process, resulting in more accurate maps with a reduced number of spurious objects. However, to the best of our knowledge, no examples of robotic systems capable of autonomously mapping an indoor crowded environment based on LiDAR SLAM

* Corresponding author

and deep learning-based people detection can be found in the present literature. In this work, we propose a mapping system based on an autonomous mobile robot equipped with a LiDAR device and a RGB camera, that can deal with the presence of people. Thanks to a deep learning approach, the position of humans is identified and a new surveying path is planned that brings the robot to scan occluded areas, so as to obtain a complete point cloud of the environment. Experimental results carried out with a wheeled mobile robot in different crowded scenarios show the applicability of the approach and the feasibility of implementation.

The organization of the paper is the following: the proposed framework is presented in Sect. 2, where the SLAM algorithm, the exploration and navigation approach and the people detection technique are described. Sect. 3 reports the experimental results and the discussion. Finally, Sect. 4 concludes the paper.

2. PROPOSED FRAMEWORK

To deal with the presence of people that can cause occlusions and consequently data gaps in the point cloud of the environment, we propose a mapping procedure characterized by the following three main steps (this process can be iterated until a people-free point cloud is obtained):

1. *Initial survey of the area.* A first tentative LiDAR-based model is acquired by the UGV that autonomously explores the surroundings using a frontier search based algorithm, and concurrently captures images of the crowded environment.

2. *People detection.* Contextually to the exploration, people identification is performed on the RBG images exploiting a Convolutional Neural Network (CNN) developed for object detection. Following an approach similar to (Ayala-Alfaro et al., 2021), we combine the bounding box provided by the CNN, the range values recorded by the LiDAR, the pose of the UGV retrieved through the SLAM algorithm, and the known transformations between LiDAR, camera and the UGV frames to estimate, with respect to a fixed frame, the position of human bounding volumes. The latter allow the developed algorithm (1) to automatically remove points corresponding to people from the model, which can result in unwanted noisy and spurious objects in the 3D map, and (2) to define the positions that the robot has to visit again.

3. *Remapping of occluded surfaces.* The UGV autonomously navigates to the positions from which people were seen, in order to map the regions of the environment previously occluded by the people. The final model is obtained by merging the point clouds acquired at different times.

An overview of the proposed framework is illustrated in Fig. 1. In the flowchart, the main blocks of the framework and the relative connections between them are explained. LiDAR data are employed in the 3D SLAM blocks to retrieve the current position of the robot and build the global map, whereas images from the camera are used to detect people. The core algorithm is *people detection* in which the map is cleaned by people and way points for the second navigation stage are computed. The position of the robot is used, together with the horizontal LiDAR scan, by a 2D graph SLAM algorithm to update the occupancy grid. The frontier exploration algorithm, indeed, needs the occupancy grid to provide way points to the global path planner during the first survey. The path computed by the global path planner is modified based on the horizontal LiDAR scan to perform obstacle avoidance and send velocity command to the robot driver. Each section of the algorithm is explained more in detail in the following.

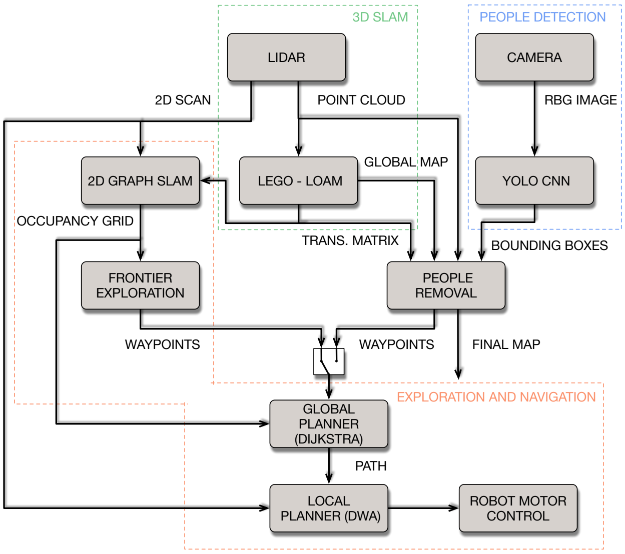


Figure 1. Flowchart of the proposed framework.

To better appreciate the performance of the proposed method and the autonomous mapping with the mobile robot, a video is available online[1].

2.1 3D SLAM algorithm

To localize the mobile robot and build a 3D reconstruction of the surroundings, a LiDAR SLAM algorithm is required. In particular, the SLAM method needs to meet the following requisites to fit with the purposes of the proposed approach: it should run in real time, provide a highly dense point cloud, and retrieve an accurate estimate of the robot position. *Light weight and Ground Optimized LiDAR Odometry and Mapping* (LeGO-LOAM), proposed in (Shan and Englot, 2018), is a ground optimized six degrees of freedom SLAM algorithm that suits these requirements, allowing also to take into account loops closing to avoid drift in the computed trajectory. Moreover, it limits the computational load usually required for scan-matching by performing feature extraction and matching between subsequent scans.

The first stage implemented in LeGO-LOAM focuses on distinguishing between ground and other objects, that are subsequently clustered, following the approach proposed in (Bogoslavskyi and Stachniss, 2016). More in detail, the point cloud is first projected onto a range image, which allows to exploit the cleanly defined neighborhood relation among points. Points that may represent the ground are identified and the remaining ones are grouped into clusters. Edge and planar features are then extracted from ground and clustered points, instead from a raw point clouds. By calculating a parameter c proportional to the differences in range measures between a point and the points in its local region, the point with high values of c are selected as edge features, and point with low values of c are

[1] `https://youtu.be/nslraeiyeUM`

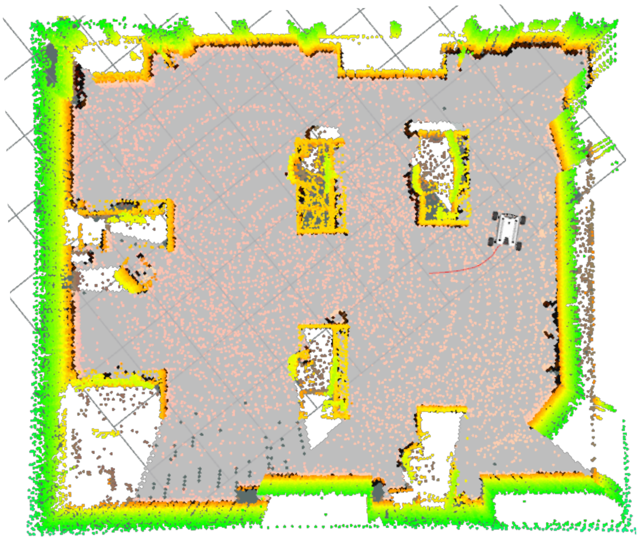


Figure 2. Occupancy grid and three dimensional map built by LeGO-LOAM. The current local path for exploration is displayed as a red line.

designated as planar features. The transformation between two subsequent scans is performed by point-to-edge and point-to-plane scan-matching. Correspondences are found only among features with the same label, thus improving the matching accuracy.

Instead of solving the non-linear expressions for the distances between edge and planar features in a unique vectors, a two-step Levenberg-Marquardt optimization method is introduced. The translation t_z along the z axis (normal to the ground) and the rotations θ_{roll} and θ_{pitch} with respect to the longitudinal and lateral axes of the mobile robot, respectively, are estimated by matching the planar features and their correspondences. On the other hand, the translations t_x and t_y on the x and y axis of the ground plane, and the rotation θ_{yaw} of the robot with respect to the ground normal are estimated using the edge feature and their correspondences.

2.2 Exploration and navigation approach

The navigation approach adopted in this paper is based on a classical path planner on a 2D flat surface, in the form of an occupancy grid, as shown in Fig. 2. An occupancy grid is similar to an image with pixel values associated to a number that indicates the likelihood the cell is occupied. If the likelihood is under a user defined threshold, the cell is marked as free and, therefore, it is accessible by the UGV. For exploring purposes, paths towards unexplored cells (marked in the map with -1 values) can be considered feasible.

The navigation framework[2] employs a pose graph SLAM algorithm to provide the above mentioned map. Graph nodes correspond to the poses of the robot, at different instants of time, whereas edges represent constraints between the nodes, as stated in (Grisetti et al., 2010). The bi-dimensional SLAM algorithm requires: (1) odometry measurement that is provided by LeGO-LOAM; and (2) perceptions of the environment, obtained from the horizontal LiDAR scan.

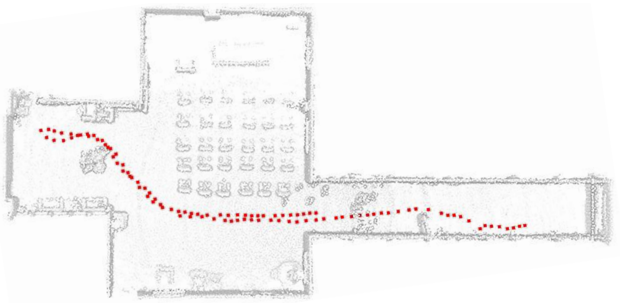


Figure 3. Trajectory of the robot performing a survey in the second test case environment.

Generally, a map of the environment must be given in advance to the robot to perform a navigation tasks. However, for autonomous mapping purposes the robot should gather information about the surrounding with a ”bootstrap” process. We define *exploration* the act of moving through an unknown environment while building a map that can be used for subsequent navigation. In this context, the concept of frontier is useful: *frontiers* are regions on the boundary between explored and unexplored space. Starting from an occupancy grid, provided by the SLAM algorithm, a method similar to edge detection in digital images is used to find the frontiers (Yamauchi, 1997). Moving to successive frontiers, the robot increases its knowledge of the world, until no more frontiers are detected and the environment can be considered explored. To provide the robot with autonomous exploration capability we used *explore_lite*[3], a ROS package that performs a greedy frontier-based exploration.

To move across a series of way points, the robot uses the ROS navigation framework, which comprises a global and a local path planner to generate a safe path (without collisions with static and dynamic obstacles) for the robot (Fig. 3). The occupancy grid is used to compute a cost function, proportional to the risk of collision. The cost function is used to weight edges into a graph: each pixel becomes a node and each adjacent nodes pair is connected by two oriented edges. The edge takes the weight from the value of the end pixel. The global planner uses the well known Dijkstra algorithm (Dijkstra, 1959) for finding the shortest paths between nodes.

In addition, since the robot is supposed to move in a unknown and dynamic environment with people moving inside it, a local planner for obstacle avoidance is also needed. For this purpose, the Dynamic Window Approach (DWA) (Fox et al., 1997) is employed, which relies on a local cost map that is generated and continuously updated using range measurements provided by LiDAR scans. For each iteration of the planner, the algorithm generates multiple tentative linear and angular velocities of the robot. For each of these velocities, the corresponding path of the robot is simulated. Paths resulting in a collision with obstacles are discarded. The remaining collision-free ones are evaluated in terms of proximity to obstacles, to the goal, and to the global path. Once the best local path is chosen, the associated velocity command is sent to the driver of the robot.

2.3 People Detection

While the process of building the map with the UGV relies on the LiDAR point cloud data only, the available methods of object detection from raw point-cloud data are faced with the

[2] http://wiki.ros.org/navigation

[3] http://wiki.ros.org/explore_lite

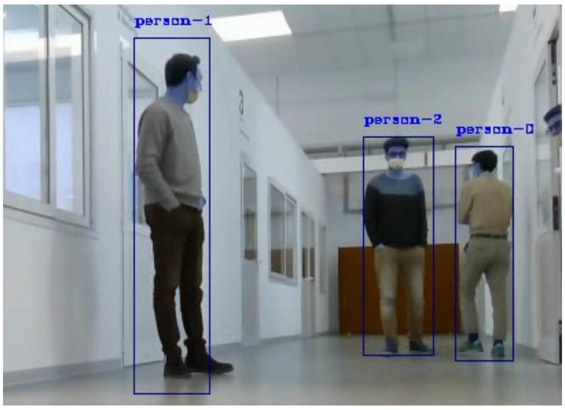


Figure 4. People detected by the YOLOv3 framework on the image acquired by the camera.

problem of the irregular structure of the data format and large search space for geometric features. This consequently reduces the performance for simultaneous processing of people detection to remove resulting occlusions and map building. To overcome this problem, we implement a method that effectively combines image and point-cloud data acquired by the camera sensor and the LiDAR scan, respectively. We exploit the image-based object detection method implemented by (Redmon and Farhadi, 2018) and trained on the COCO dataset, which is popularly known as YOLOv3 (Fig. 4). This CNN allows the processing of high frame rate and guarantees accuracy and robustness to objects scaling on image frame, thus it is particularly suited for our application. With this framework, we can retrieve people position in real-time from the point cloud data by projecting predicted points of each bounding box from the camera frame to the LiDAR frame. This approach follows two main steps: (1) obtaining the transformation function of the LiDAR to the camera frame, and (2) locating the 3D points corresponding to the vertices of a 2D bounding box in the image frame.

To obtain the LiDAR-to-camera transformation, the knowledge of the extrinsic parameters of the sensors is required. These parameters express the relative pose of the LiDAR and camera sensor, which can be achieved through a calibration process. Thanks to OpenCV Perspective-n-Point (PnP) pose computation, the rotation and translation transform between the sensors is determined from a set of manually picked corresponding 2D and 3D points in the individual sensor frame, in addition to the camera intrinsic matrix and the distortion coefficients. With the LiDAR-camera extrinsic parameters, a perspective projection is computed to obtain the corresponding 2D projection of the 3D point cloud data. People's positions in the LiDAR reference frame are retrieved by taking the corresponding 3D coordinates of the projected LiDAR point closest to the bounding box center (Fig. 5).

Whenever a detection occurs, the most recent available pose of the robot, with respect to the fixed reference frame is used to transform people coordinates from the LiDAR to the fixed reference frame. The so retrieved positions are stored in a buffer and used at the end of the exploration stage to clean the map. When no more frontiers are detected (or after a user defined period of time) the exploration is considered complete and the cleaning process starts. People positions are employed as centers of three dimensional bounding boxes, of fixed dimension, in the global map built by LeGO-LOAM. Thus, points inside boxes are sequentially removed from the map. After this operation, the second navigation stage begins and the robot comes back to previous positions, from which people were seen, to fill possible occlusions and data gaps in the point cloud.

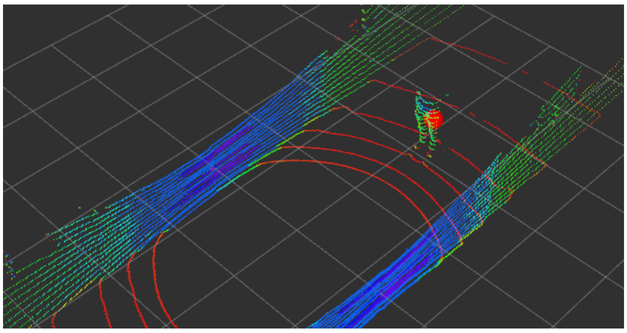


Figure 5. People positions in the LiDAR reference frame. The red marker represent the center of the three dimensional bounding box.

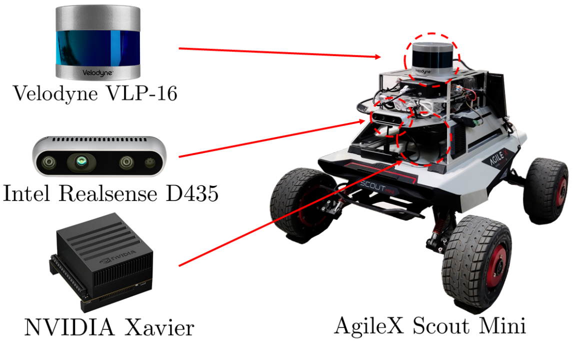
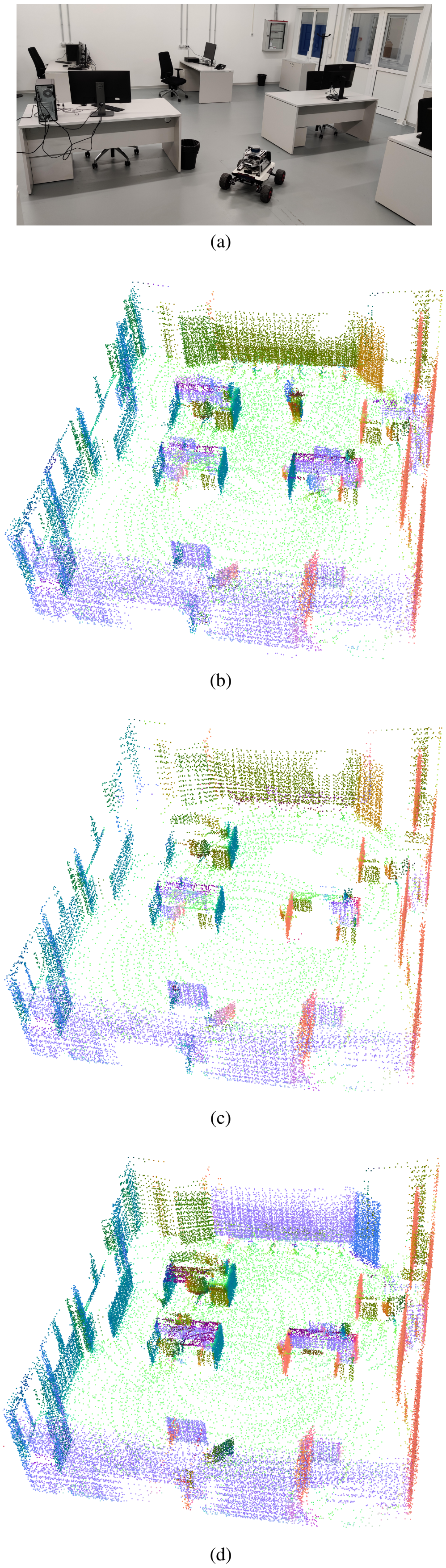


Figure 6. Mobile robotic platform equipped with 3D laser scanner, RGB camera and onboard computer.

3. EXPERIMENTAL RESULTS

Experimental tests were carried out implementing the proposed framework on a Scout Mini UGV by AgileX Robotics. This mobile robot, shown in Fig. 6, is equipped with a Velodyne VLP-16 laser scanner, an Intel Realsense D435 stereo depth camera, and a NVIDIA Xavier computer. The laser measurement range is up to 100 m with a precision of ±3 cm. Its vertical field of view is 30° (±15°), and the horizontal field of view is 360°. Moreover, it provides a vertical angular resolution of 2° and a horizontal resolution of 0.2°, as the rotation rate is set to 10 Hz. The Intel Realsense camera has a maximum RGB resolution of 1920 x 1080 and a lower maximum range (10 m) with respect to that of the LiDAR. The camera depth measurements resulted to be noisy, thus they were not used to detect people's position in this work. Finally, the NVIDIA Jetson AGX Xavier board is equipped with a graphics card powered by 512 Tensor cores, 32 GB of RAM and a 8-Core ARM v8.2 64-bit CPU, and runs Ubuntu 18.04 with the Robot Operating System (ROS Melodic).

The proposed method was tested to map two areas of a building that hosts the robotics laboratory, within the scientific campus of the University of Udine (Italy). The two test cases that were surveyed in the presence of people consist in (1) a single closed office (Fig. 7(a)), and (2) a conference room comprising a long and narrow corridor (Fig. 8(a)).

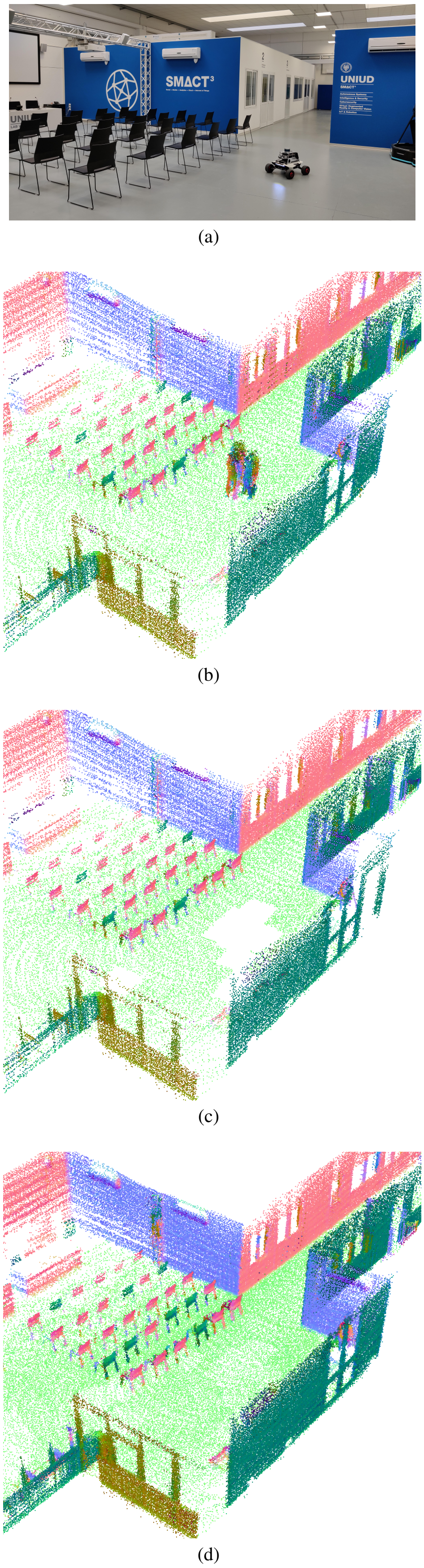

(a)

(b)

(c)

(d)

Figure 7. Test case (1): scanned environment (a); people detected (b); point cloud with data gaps (c); final point cloud (d).


(a)

(b)

(c)

(d)

Figure 8. Test case (2): scanned environment (a); people detected (b); point cloud with data gaps (c); final point cloud (d).

The first environment is characterized by the presence of tables, chairs and furniture, representing static obstacles for the UGV. Moreover, during the survey, a person was moving inside the office. The robot took approximately two minutes to perform the first exploration and to remap the previously occluded areas. Fig. 7(b)-7(d) illustrate the result for test case (1), where the first point cloud is characterized by noisy and spurious points associated to the presence of people, that are automatically removed, as shown in Fig. 7(c). On the other hand, the removal causes square holes in the point cloud that are filled during the subsequent survey, as it can be seen in Fig. 7(d).

The second test case was performed in a wider and more unstructured space, with half of the area occluded by chairs. As shown in Fig. 3, the robot succeeded in finding a safe path to explore even this challenging environment, with three moving people. The overall required time to perform the mapping operation was about five minutes. Final results obtained with the proposed workflow are presented in Fig. 8(b)-8(d). Please note that all the people were correctly identified in the images by the CNN and removed from the point cloud (Fig. 8(c)). However, a more robust implementation of the position retrieval method in the LiDAR frame should be used to avoid the incorrect removal of points not belonging to people. This may be caused by the wrong positioning of the volume bounding boxes in the fixed reference frame.

4. CONCLUSION

In this paper, we presented a mapping system based on a mobile robot equipped with a LiDAR device and a RGB camera, that can deal with the presence of people. Thanks to a deep learning approach, the position of humans is identified and a new surveying path is planned that brings the robot to scan occluded areas, so as to obtain a complete point cloud of the environment. Experimental results were performed with a wheeled mobile robot in different crowded scenarios, showing the applicability of the proposed approach to perform an autonomous survey avoiding occlusions and removing noisy and spurious objects caused by people presence from the map.

In future developments of this work, we plan to further investigate the problem of mapping of crowded environments by means of mobile robots. In particular, we will consider alternative SLAM algorithms, as well as other path planning strategies for the navigation of the robot in an unknown environment. Furthermore, we will test the proposed framework in more complex scenarios, such as outdoor areas and cluttered environments.

ACKNOWLEDGMENTS

This research has been developed within the Laboratory for BigData, IoT, Cyber Security (LABIC) funded by Friuli Venezia Giulia region, and the Laboratory for Artificial Intelligence for Human-Robot Collaboration (AI4HRC) funded by Fondazione Friuli. The first author acknowledges support from the National Ph.D. Programme in Artificial Intelligence (PhD-AI.it) of University of Naples Federico II (Naples, Italy).

REFERENCES

Adán, A., Quintana, B., Prieto, S. A., 2019. Autonomous mobile scanning systems for the digitization of buildings: A review. *Remote Sensing*, 11(3), 306.

Asadi, K., Ramshankar, H., Pullagurla, H., Bhandare, A., Shanbhag, S., Mehta, P., Kundu, S., Han, K., Lobaton, E., Wu, T., 2018. Building an integrated mobile robotic system for real-time applications in construction. *arXiv preprint arXiv:1803.01745.*

Ayala-Alfaro, V., Vilchis-Mar, J., Correa-Tome, F., Ramirez-Paredes, J., 2021. Automatic Mapping with Obstacle Identification for Indoor Human Mobility Assessment. *arXiv preprint arXiv:2111.12690.*

Bogoslavskyi, I., Stachniss, C., 2016. Fast range image-based segmentation of sparse 3d laser scans for online operation. *IEEE/RSJ Int. Conf. on Int. Rob. and Syst.*, IEEE, 163–169.

Dijkstra, E. W., 1959. A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1), 269–271.

Fox, D., Burgard, W., Thrun, S., 1997. The dynamic window approach to collision avoidance. *IEEE Robotics & Automation Magazine*, 4(1), 23–33.

Giusti, A., Magnago, V., Siegele, D., Terzer, M., Follini, C., Garbin, S., Marcher, C., Steiner, D., Schweigkofler, A., Riedl, M., 2021. BALTO: A BIM-Integrated Mobile Robot Manipulator for Precise and Autonomous Disinfection in Buildings against COVID-19. *2021 IEEE 17th Int. Conf. on Automation Science and Engineering (CASE)*, 1730–1737.

Grisetti, G., Kümmerle, R., Stachniss, C., Burgard, W., 2010. A tutorial on graph-based SLAM. *IEEE Int. Transp. Syst. Mag.*, 2(4), 31–43.

Hähnel, D., Schulz, D., Burgard, W., 2003. Mobile robot mapping in populated environments. *Adv. Rob.*, 17(7), 579–597.

Maset, E., Cucchiaro, S., Cazorzi, F., Crosilla, F., Fusiello, A., Beinat, A., 2021. Investigating the performance of a handheld mobile mapping system in different outdoor scenarios. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, XLIII-B1-2021, 103–109.

Maset, E., Scalera, L., Beinat, A., Cazorzi, F., Crosilla, F., Fusiello, A., Gasparetto, A., 2022. Preliminary comparison between handheld and mobile robotic mapping systems. *Proc. of I4SDG Workshop 2021*, Springer International Publishing, Cham, 290–298.

Park, D., Hwang, S., 2020. Autonomous Indoor Scanning System Collecting Spatial and Environmental Data for Efficient Indoor Monitoring and Control. *Processes*, 8(9), 1133.

Rausch, C., Haas, C., 2021. Automated shape and pose updating of building information model elements from 3D point clouds. *Automation in Construction*, 124, 103561.

Redmon, J., Farhadi, A., 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767.*

Shan, T., Englot, B., 2018. LeGO-LOAM: Lightweight and ground-optimized lidar odometry and mapping on variable terrain. *2018 IEEE/RSJ Int. Conf. on Int. Rob. and Syst.*, IEEE, 4758–4765.

Yamauchi, B., 1997. A frontier-based approach for autonomous exploration. *IEEE Int. Symp. on Comp. Intell. in Rob. and Autom.*, IEEE, 146–151.

Zhang, Y., Tian, G., Shao, X., Liu, S., Zhang, M., Duan, P., 2021. Building Metric-Topological Map to Efficient Object Search for Mobile Robot. *IEEE Trans. on Ind. Electr.*