Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it Bonn-Aachen International Center for Information Technology

R&D Project Proposal

Semantic Mapping for Enhanced Localization in Indoor Environments

Allen Isaac Jose

Supervised by

Prof. Dr.-Ing. Sebastian Houben
Deebul Nair, M.Sc.

January 2023

1 Introduction

1.1 Topic of This R&D Project

- The number of Autonomous Mobile Robots(AMR) being deployed into the home and industrial environments is increasing day by day. This is due to the fact that they can be reliable in performing the task assigned to them redundantly in a robust manner.

- For this to happen, the robot should have an idea or knowledge of the structure of the environment so that it can plan its actions accordingly. This problem in the field of robotics is known as mapping. Mapping is the process of creating a virtual model of the real environment that can be used by robots for various applications such as localization, navigation, and planning.

- Another important requirement is the ability of the robot to be able to locate itself within the map it has created. This problem is known as localization. It has to match the observation from the sensors with the structure or objects within the available map to identify correspondences and calculate its pose with respect to the map of the environment.

- When the ground truth pose/location of the robot is given, the task of mapping could be done directly and similarly, when the map of the environment is provided, the task of localization can be easily achieved. But in reality, when the robot needs to explore a new environment, it has to perform both localization and mapping at the same time. This well-known problem is known as Simultaneous Localization and Mapping(SLAM).

- In SLAM, we need to estimate the pose and generate a map from the observations simultaneously. The main challenge is that the action model and observation model of the robots is not perfect. That is, there is uncertainty in the final pose of a robot given a command for action and similarly, there is uncertainty associated with each sensor when observing an environment. These uncertainties of sensors need to be taken care of, especially in the case of sensor fusion.

- Previously, sensors like laser range finders, LIDAR, and cameras were popularly used in this field. But now, due to the availability of high-quality RGB-D cameras, a dense reconstruction of the environment was made possible with the point cloud operations such as Iterative Closest Point(ICP).
- The emergence of deep learning methods helps in detecting, classifying and segmenting objects in the environment, either from images or from the point clouds. Prior to deep learning models, the focus was to generate a 2D/3D representation of the environment during mapping and extracting features from the map during the process of localization. Now, the advancements in deep learning help us to incorporate semantic information of the objects into the map representation rather than just storing raw pixels or point clouds.
- In this project, we aim to perform an exploratory study to identify the failure scenarios of the existing 2D indoor-SLAM approaches in ROS and the performance improvement in localization by incorporating semantic information into the maps.

1.2 Relevance of This R&D Project

- Incorporating the semantic information can help to:
  - Improved localization in the indoor environment.
  - Reduce the memory required to store the map of large environments as raw sensor data can be replaced by semantic geometric structures.
  - Removing dynamic objects from the environment such as humans.
  - Update the map in a highly dynamic environment.
  - Provide a better understanding of the scene which can aid in the task planning process.
  - Such a semantic map can further aid the generation of positional relationships between the objects.
  - Decomposing a bigger map into separate rooms and thereby allowing better modularity.

- The traditional methods that extract features from the 2D or 3D environment for correspondence matching may take more time for localization as the size of the map increases. When using semantic knowledge-based maps, the overlapping between the observation and the virtual map based on semantically labelled objects can help in faster localization even if the object pose is changed by a small quantity.

- The normal 3D representation of an environment is a dense point cloud which is stored as a data structure depending on the type of map. Clustering and giving semantic meaning to a group of points can help to replace the point cloud with quadrics like cuboids or use an object mesh which can significantly reduce the memory required to store the map [30], [23]. This can also fasten the localization process as compared to traditional 3D methods as we only need to find the overlapping between known objects rather than matching the point clouds which can be erroneous in certain sensor spoofing situations.

- Oftentimes when the robot is performing the mapping operation, there can be chances that humans may appear in some of the frames. This causes an effect called ghosting where multiple copies of the same human can appear at different locations within the generated map as the person moves continuously. But on the map, such dynamic objects should not be there. Through semantic detection of humans, we can remove the point clouds or map representation associated with humans.

- Lifelong SLAM is a capability of a system to detect the changes in the environment and hence update the map it has created for the long term. Without semantic information, they had to use feature-based change detection algorithms. Whereas in semantic maps, the clustering of point clouds into a segment can help in easier tracking of the objects. And when an object changes its pose or is removed or added to the environment, corresponding changes can be easily done to the semantic maps.

- The availability of semantic maps can help the robot understand the location where an object is most likely to be present. For example, kitchen-related equipment can be found in the kitchen. Also, the knowledge of the position and pose of the object, the robot is searching for, can help in the efficient generation of a sequence of plans to perform the task. In the case of navigation planning, semantic maps can help in choosing a safer path as it knows the risk elements within each possible route.

- The semantic information can help to derive the positional relationship between the objects such as the bottle being on the table, the mug being inside the cupboard, the bed being close to the door, the pendant light being located above the table.

- The indoor environment to be mapped may consist of different sections or rooms like office rooms, kitchens, hallways, and large halls. If the map exists as a singular one, with the reconstruction of all the rooms together, it may be difficult for managing and updating a piece of single room information within it later. With the semantics of walls, floors, and entry-exit points(doors), the map can be decomposed into separate sections and can be connected topologically with the doors as entry or exit points. This enables the easier remapping of a particular room without affecting other parts of the environment. Also, it helps to assign multiple maps of the same room, each with different semantic information, with the particular node that represents the room.

- Some of the industries that are benefiting from such indoor semantic SLAM approaches are:

  - Augmented Reality(AR) industry
  - Autonomous Mobile Robots(AMR) in industrial applications
  - Human Support Robots(HSR) in home applications

2 Related Work

2.1 Survey of Related Work

- In the initial stages of SLAM, a 2D LIDAR or laser range finder was used to identify the distance to the obstacle and generate an occupancy grid map containing free spaces and obstacles. Some of them are Hector SLAM[16], GMapping[11], Karto SLAM[17].

- Later when cameras were made available, visual features were able to be extracted and used in SLAM for operations such as loop closure. They were collectively called as visual SLAM methods. Along with it, the availability of point clouds from RGB-D cameras helped in the dense reconstruction of the environments. Some of them are RatSLAM[20], MonoSLAM[4], Parallel Tracking and Mapping(PTAM)[15], Dense Tracking and Mapping(DTAM)[22], RGB-D Mapping[12], Real-Time Appearance-Based Mapping(RTAB-Map)[18], Large-Scale Direct Monocular SLAM(LSD-SLAM)[7], Dense Piecewise Planar Tracking and Tracking(DPTTAM)[3], Oriented FAST and Rotated BRIEF(ORB-SLAM)[21], ElasticFusion[29], Google Cartographer[13], Direct Sparse Odometry(DSO)[8].

- Advancements in deep learning further accelerated this map creation process by helping in object classification and object segmentation either from raw RGB images or directly on point clouds. This helped to include semantic information in the map. Some of the examples are SLAM++[27], SemanticFusion[19], MaskFusion[26], CDSFusion[28], and Kimera[25]. In certain methods, the object is represented using cuboids[30] or using quadrics[23] where a prior object model is not required. Google has developed Objectron[1] which is able to form a 3D bounding box on 2D images with the knowledge of the prior trained info and is able to track its pose efficiently even on a mobile device.

- Kalman filter is a filter belonging to the extension of the Bayes filter which was used to combine noisy observations to derive the best estimate of the observation for localization. But the Kalman filter assumes that the noises are

Gaussian in nature or in other words uni-modal distribution. This led to the formation of an Extended Kalman Filter(EKF) where the non-linearities are approximated using the Taylor Series. But still, the real environment is highly non-linear in nature which limits the convergence performance of the Kalman filter family. Hence particle filter-based localization[10][2] was proposed which can represent multi-modal distributions at the cost of increased computation power. The current Adaptive Monte-Carlo Localization(AMCL)[9] package in ROS is built based on the Rao-Blackwellized version of the particle filter[11].

- In dense RGB-D reconstructed maps without any semantic information, localization is performed by extracting features from the observation and identifying the 2D-3D correspondences within the map it has created or in other words, feature matching[14], [31], [24]. But the problem is, as the size of the map increases, the time required to identify the correspondence increases and may not be optimal for a real-time operation. Also, the assumption of the environment being static may not be practical in real-life scenarios.

- In the work by Yang et al.[30], they use RGB images from monocular camera to fit cuboids to the objects in the environment. It is built on top of ORB-SLAM2 and modified the bundle adjustment process such that the camera pose estimation and object pose estimation are done parallelly which can benefit the map making process. A YOLO object detection model is used to obtain the bounding box in RGB images and proposals for cuboids are generated with respect to the edges of this bounding box along with taking advantage of vanishing point-based geometry. The advantage is that this approach doesn't need a prior object model as compared to deep learning-based 3D cuboid creation methods. The optimization of these cuboid are based on an objective function and through multi view optimization. Also, the dynamic objects can be detected and tracked as compared to older approaches like slam++ where dynamic object is considered as an outlier[30].

- In the work by Dengler et al.[5], a semantic map of the environment is mapped onto a 2D grid map where each object is represented by a fitting polygon. This is stored along with the label and 3D point cloud information of that object.

For operations such as navigation planning to reach the object, a 2D semantic map is sufficient. And for manipulation of the object, it then utilizes the point cloud information of that object. Also, an existence likelihood probability is calculated for each object which can be used to continuously update the map in case of a dynamic environment. Object instance segmentation is done using mask R-CNN and geometric segmentation is done by performing segmentation on the point cloud directly. The system was capable of performing online mapping in a Toyota HSR robot[5].

- Kimera[25] is a semantic SLAM framework with the dense reconstruction of the environment using fast inertial measurement data from IMU and stereo images. It consists of mainly 4 components namely, a visual-inertial odometry(VIO) module for state estimation, a robust pose graph optimizer(RPGO), a fast 3D local mesher and a slower 3D global dense semantic reconstruction module. The framework is able to execute in real-time on the CPU by utilising multi-threading. The Visual-Inertial Odometry Module performs state estimation on the frontend using the IMU data between consecutive frames and is verified using feature detection, stereo matching, and geometric verification from stereo pair. On the backend, a factor graph is formed which is optimised by the RPGO module using loop closures detection in GTSAM. The fast mesh calculation in each keyframe is done to avoid obstacle collision during navigation. The multi-frame meshes are fused together to form a global mesh and semantic labels are assigned to it[25].

- In the floor plan-based localization presented by Zimmerman et al.[32], the prior available CAD-based floor plan is enriched with semantic cues of the objects in the environment and is used for localization. They state that due to the architectural similarity in the rooms within an indoor environment, floor plans alone are not enough for localization. For Monte-Carlo localization, they are providing multimodal sensor information when they are available. Odometry data for the motion model, 2D LIDAR data for the beam-end model, and object detection from RGB images for the semantic visibility model. YOLOv5s is used for object detection. For efficiently initializing the particles, a nearest-neighbour classifier is also used to identify the room based on encoded information on the objects that are usually seen in a particular room. To each free cell within the occupancy grid map, the object label it can observe without any obstruction along with the direction vector to the object is stored. A cosine similarity distance is used to compare the particle's pose and the pose of cells in the created semantic map to localize the robot[32].

2.2 Limitation and Deficits in the State of the Art

- In the work by Yang et al.[30], the algorithm is able to form 3D cuboids only on the objects on the ground. This was done to make the process computationally feasible in proposal sampling for 3D cuboids by taking advantage of the knowledge of the height of the camera. The 2D bounding box generated by YOLO would affect the 3D cuboid accuracy since the proposals for 3D cuboids are generated by utilizing the bounding box edges. Using an RGB-D camera instead of an RGB camera may help rectify this problem with the help of point cloud depth information and clustering.

- In the work by Dengler et al.[5], the localization is done with a traditional gmapping-based map and AMCL-based localization approach. That is, the objects are not considered as an observation while localizing within the environment. The potential of the semantic information gathered is not fully utilized as it is used only for planning the manipulation task. Also, the wrong association of point clouds with another instance of the object led to the extension of the object into an unintended shape[5].

- The time taken for pose graph optimization in Kimera[25] increases as the size of the environment increases. Currently, the method is tested only on the EUROC V101 dataset which is a single-room indoor environment. Even though they mentioned the mapping process in SLAM, they failed to mention about relocalizing in such an environment. The dense stereo method finds it difficult to estimate the depth of textureless components in the environment like walls, floors, and shelves causing to have both geometric and semantic errors in the reconstructed map[25].

- In the floor plan-based localization presented by Zimmerman et al.[32], the semantic cues are to be added manually using a tool they have developed. It assumes that there won't be any drastic change in the position of objects in the environment which cannot always be true. Also, manual effort is needed in semantic map creation and for updating, if there are drastic changes in the environment. Furthermore, only objects on the ground can be used in the observation model as it uses the range information from the 2D LIDAR sensor to estimate the semantic visibility of objects through ray tracing. Also, the static structure of the room is not at all taken into consideration which would have been a piece of useful semantic information[32].

- Currently, the map of the indoor environment is reconstructed into a single fused one. Therefore, updating the information of a single room is difficult as the operation needs to be performed on the entire map module. Decomposition of a large map into smaller rooms based on topological structure helps in better scalability and provides easier modification strategies.

- Even though particle filter-based localization is widely accepted among the robotics community, there are still some challenges existing in this domain which are mentioned by Elfring et al.[6]. Five of them are degeneracy problems, sample impoverishment, particle filter divergence, selecting the importance density, and real-time execution. Some of the solutions are suggested by the author itself but need to be implemented and tested in a simulation environment[6].

- A major topic which is missing in most of the proposed SLAM approaches is how to relocalize in an environment that has already been mapped. While some still use the Adaptive Monte-Carlo Localization(AMCL), they are not utilising the semantics and pose information of the object in the observation model. Furthermore, there is a question arising on how to efficiently perform particle filter localization in a 3D environment. The observation space exponentially increases and also needs to deal with how to initialize the particle and the count of particles.

3 Problem Statement

- The proposed project aims to identify the scenarios in which GMapping + AMCL-based localization fails, analyze the reason for the failure and utilise a semantic SLAM approach to overcome the situation.
- The proposed project seeks to address the challenges in semantic SLAM such as how to relocalize efficiently in an environment that has already been mapped using the semantic SLAM approach and how to efficiently encode dense semantic point cloud information for faster processing and easier storage of the map.
- For that, initially we need to demonstrate the situations in which the traditional 2D SLAM approach based on GMapping and AMCL in ROS fails. These scenarios are to be created in the Gazebo simulation environment. The next task is to implement at least one of the existing semantic SLAM approaches to tackle these failure situations in Gazebo.
- The final task is to integrate the algorithm into the Toyota HSR robot and test the semantic mapping and localization task in a real-world indoor environment. It will also give an insight into the execution performance difference between simulated and real-world environments which can aid for the improvements in the algorithm.
- Since localization is one of the important research topics in this project, Time taken to localize, uncertainty or error in the estimated position and pose would be the evaluation metrics. In the case of semantic mapping, the Intersection over Union(IoU) can be a useful comparison feature. True positive rate and false positive rate can also be evaluated in the case of labels being predicted for the objects in the environment.
- Thus the main two outcomes of the project are the generation of a semantic map through SLAM and then relocalizing in such an environment.
- This work will also provide a brief idea of how the framework of semantic SLAM looks like in the existing approaches.

4 Project Plan

4.1 Work Packages

WP1 Literature Review

- T1.1 Survey the existing indoor mapping approaches(example: GMapping, HectorSLAM, KartoSLAM, Google cartographer etc.)
- T1.2 Survey the existing localization approaches(example: Probabilistic localization approaches like Adaptive Monte Carlo Localization(AMCL))
- T1.3 Survey the existing semantic SLAM approaches(example: KimeraSemantics, QuadricSLAM etc.)

WP2 Creation of Gazebo environment models where AMCL-based localization fails in GMapping-based maps.

WP3 Evaluation of the existing GMapping + AMCL based localization

- T3.1 Experimentation by running the GMapping + AMCL-based localization approach on the created Gazebo environments.
- T3.2 Identifying and analyzing the possible reasons for the failure.

WP4 Identifying the semantic SLAM approach

- T4.1 Compare the underlying framework of various existing semantic SLAM approaches.
- T4.2 Collect the codebase for the selected semantic SLAM approaches and generate a ROS-compatible modified version to meet the needs of indoor SLAM.

WP5 Evaluation of the semantic SLAM approach

- T5.1 Experimentation by running the formulated semantic SLAM approach on the created Gazebo environments
- T5.2 Evaluate results based on localization metrics such as time taken to localize, and accuracy of the estimated pose with ground truth. Also

semantic mapping metrics such as Intersection over Union (IoU), and object label accuracy in terms of true and false positive rates.

T5.3 Integrate and experiment with the semantic SLAM approach in Toyota HSR robot in an indoor environment. (Desired deliverable)

WP6 Project Report

T6.1 Prepare an initial draft of the report.

T6.2 Collect peer review and expert review comments.

T6.3 Fix the review comments.

T6.4 Submission of the final draft of the report.

4.2 Milestones

M1 Literature review completed and best practice identified.

M2 Completion of Gazebo environment creation.

M3 Experimentation and evaluation of GMapping + AMCL approach in Gazebo environment.

M4 Completion of ROS-compatible semantic SLAM algorithm.

M5 Experimentation and evaluation of semantic SLAM approach in Gazebo environment.

M6 Experimentation and evaluation of semantic SLAM approach in Toyota HSR robot. (Desired deliverable)

M7 Report submission

4.3 Project Schedule

WORK PACKAGES	START DATE	END DATE	DAYS	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept
Literature Review	**01-02-2023**	**31-05-2023**	**33**								
Survey the existing indoor mapping approaches	01-02-2023	15-02-2023	5								
Survey the existing indoor localization approaches	16-02-2023	23-02-2023	4								
Survey the existing semantic SLAM approaches	25-03-2023	31-05-2023	24								
Gazebo environment creation	**01-02-2023**	**05-03-2023**	**19**								
Familiarize with ROS mapping algorithms	01-02-2023	15-02-2023	6								
Familiarize with ROS localization algorithms	16-02-2023	23-02-2023	3								
Identifying AMCL localization failure scenarios	16-02-2023	05-03-2023	6								
Gazebo environment creation	23-02-2023	05-03-2023	4								
Deficits of existing methods	**05-03-2023**	**20-03-2023**	**11**								
Experimentation and evaluation of AMCL localization in Gazebo environments	05-03-2023	20-03-2023	6								
Identifying and analyzing the possible failure reasons	05-03-2023	20-03-2023	5								
Identify semantic SLAM algorithm	**25-03-2023**	**31-07-2023**	**69**								
Compare the underlying framework of various existing semantic SLAM approaches.	25-03-2023	31-05-2023	24								
Code implementation debugging and testing on given dataset	25-03-2023	31-05-2023	24								
Improvements in the existing semantic SLAM algorithm to suit the localization process and make it ROS compatible	01-06-2023	31-07-2023	21								
Evaluation of the semantic SLAM approach	**01-06-2023**	**30-09-2023**	**54**								
Experimentation and evaluation of Semantic SLAM algorithm in Gazebo environments	01-06-2023	31-07-2023	21								
Integrating and experimenting the semantic SLAM approach in Toyota HSR robot	01-08-2023	31-08-2023	11								
Improvements in the code based on the performance	01-08-2023	30-09-2023	22								
Project report	**01-02-2023**	**30-09-2023**	**91**								
Introduction, State of the art	01-02-2023	31-05-2023	38								
Proposed methodology, Implementation	01-06-2023	31-07-2023	21								
Experimentation, evaluation results	01-07-2023	31-08-2023	22								
Conclusion, refining the report	01-09-2023	30-09-2023	10								

Figure 1: Project timeline

4.4 Deliverables

Minimum Viable

- Literature review in the occupancy grid mapping strategies.
- Literature review on the problems arising in AMCL-based localization.
- Gazebo environments to simulate GMapping + AMCL failure situations.
- Final draft of the report.

Expected

- All minimum viable deliverables.
- Implementing at least one of the existing semantic SLAM approaches in Gazebo simulation to overcome the failure situations.
- ROS compatible semantic SLAM algorithm code.
- Experimental results based on evaluation metrics.

Desired

- All expected deliverables.
- Implementing and executing the above semantic SLAM algorithm in the Toyota HSR robot.

References

[1] Adel Ahmadyan, Tingbo Hou, Jianing Wei, Liangkai Zhang, Artsiom Ablavatski, and Matthias Grundmann. Instant 3D Object Tracking with Applications in Augmented Reality. *ArXiv*, abs/2006.13194, 2020.

[2] M. Sanjeev Arulampalam, Simon Maskell, Neil J. Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50:174–188, 2002.

[3] Alejo Concha and Javier Civera. DPPTAM: Dense piecewise planar tracking and mapping from a monocular sequence. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5686–5693, Sep. 2015. doi: 10.1109/IROS.2015.7354184.

[4] Andrew J. Davison, Ian D. Reid, Nicholas D. Molton, and Olivier Stasse. MonoSLAM: Real-Time Single Camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1052–1067, June 2007. ISSN 1939-3539. doi: 10.1109/TPAMI.2007.1049.

[5] Nils Dengler, Tobias Zaenker, Francesco Verdoja, and Maren Bennewitz. Online Object-Oriented Semantic Mapping and Map Updating. In *European Conference on Mobile Robots (ECMR)*, pages 1–7, 2021. doi: 10.1109/ECMR50962.2021.9568817.

[6] Jos Elfring, Elena Torta, and René van de Molengraft. Particle Filters: A Hands-On Tutorial. *Sensors*, 21(2), 2021. ISSN 1424-8220. doi: 10.3390/s21020438. URL `https://www.mdpi.com/1424-8220/21/2/438`.

[7] Jakob Engel, Thomas Schöps, and Daniel Cremers. LSD-SLAM: Large-Scale Direct Monocular SLAM. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV*, pages 834–849, Cham, 2014. Springer International Publishing. ISBN 978-3-319-10605-2.

[8] Jakob Engel, Vladlen Koltun, and Daniel Cremers. Direct Sparse Odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(3):611–625, March 2018. ISSN 1939-3539. doi: 10.1109/TPAMI.2017.2658577.

[9] Dieter Fox. Adapting the Sample Size in Particle Filters Through KLD-Sampling. *The International Journal of Robotics Research*, 22:1003 – 985, 2003.

[10] N.J. Gordon. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F (Radar and Signal Processing)*, 140:107–113(6), April 1993. ISSN 0956-375X. URL `https://digital-library.theiet.org/content/journals/10.1049/ip-f-2.1993.0015`.

[11] Giorgio Grisetti, Cyrill Stachniss, and Wolfram Burgard. Improved Techniques for Grid Mapping With Rao-Blackwellized Particle Filters. *IEEE Transactions on Robotics*, 23(1):34–46, Feb 2007. ISSN 1941-0468. doi: 10.1109/TRO.2006.889486.

[12] Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, and Dieter Fox. *RGB-D Mapping: Using Depth Cameras for Dense 3D Modeling of Indoor Environments*, pages 477–491. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. ISBN 978-3-642-28572-1. doi: 10.1007/978-3-642-28572-1_33. URL `https://doi.org/10.1007/978-3-642-28572-1_33`.

[13] Wolfgang Hess, Damon Kohler, Holger Rapp, and Daniel Andor. Real-time loop closure in 2D LIDAR SLAM. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1271–1278, May 2016. doi: 10.1109/ICRA.2016.7487258.

[14] Carlos Jaramillo, Ivan Dryanovski, Roberto G. Valenti, and Jizhong Xiao. 6-DoF pose localization in 3D point-cloud dense maps using a monocular camera. In *IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 1747–1752, Dec 2013. doi: 10.1109/ROBIO.2013.6739720.

[15] Georg Klein and David Murray. Parallel Tracking and Mapping for Small AR Workspaces. In *6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 225–234, Nov 2007. doi: 10.1109/ISMAR.2007.4538852.

[16] Stefan Kohlbrecher, Oskar von Stryk, Johannes Meyer, and Uwe Klingauf. A flexible and scalable SLAM system with full 3D motion estimation. In *IEEE International Symposium on Safety, Security, and Rescue Robotics*, pages 155–160, Nov 2011. doi: 10.1109/SSRR.2011.6106777.

[17] Kurt Konolige, Giorgio Grisetti, Rainer Kümmerle, Wolfram Burgard, Benson Limketkai, and Regis Vincent. Efficient Sparse Pose Adjustment for 2D mapping. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 22–29, Oct 2010. doi: 10.1109/IROS.2010.5649043.

[18] Mathieu Labbé and François Michaud. Online global loop closure detection for large-scale multi-session graph-based SLAM. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2661–2666, Sep. 2014. doi: 10.1109/IROS.2014.6942926.

[19] John McCormac, Ankur Handa, Andrew Davison, and Stefan Leutenegger. SemanticFusion: Dense 3D semantic mapping with convolutional neural networks. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 4628–4635, May 2017. doi: 10.1109/ICRA.2017.7989538.

[20] M.J. Milford, G.F. Wyeth, and D. Prasser. RatSLAM: a hippocampal model for simultaneous localization and mapping. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04.*, volume 1, pages 403–408 Vol.1, April 2004. doi: 10.1109/ROBOT.2004.1307183.

[21] Raúl Mur-Artal, J. M. M. Montiel, and Juan D. Tardós. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Transactions on Robotics*, 31(5):1147–1163, Oct 2015. ISSN 1941-0468. doi: 10.1109/TRO.2015.2463671.

[22] Richard A. Newcombe, Steven J. Lovegrove, and Andrew J. Davison. DTAM: Dense tracking and mapping in real-time. In *International Conference on Computer Vision*, pages 2320–2327, Nov 2011. doi: 10.1109/ICCV.2011.6126513.

[23] Lachlan Nicholson, Michael Milford, and Niko Sünderhauf. QuadricSLAM: Dual Quadrics from Object Detections as Landmarks in Object-oriented SLAM, 2018. URL `https://arxiv.org/abs/1804.04011`.

[24] Michael Paton and Jana Kosecka. Adaptive RGB-D Localization. In *Ninth Conference on Computer and Robot Vision*, pages 24–31, May 2012. doi: 10.1109/CRV.2012.11.

[25] Antoni Rosinol, Marcus Abate, Yun Chang, and Luca Carlone. Kimera: an Open-Source Library for Real-Time Metric-Semantic Localization and Mapping. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1689–1696, May 2020. doi: 10.1109/ICRA40945.2020.9196885.

[26] Martin Runz, Maud Buffier, and Lourdes Agapito. MaskFusion: Real-Time Recognition, Tracking and Reconstruction of Multiple Moving Objects. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 10–20, Oct 2018. doi: 10.1109/ISMAR.2018.00024.

[27] Renato F. Salas-Moreno, Richard A. Newcombe, Hauke Strasdat, Paul H.J. Kelly, and Andrew J. Davison. SLAM++: Simultaneous Localisation and Mapping at the Level of Objects. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1352–1359, June 2013. doi: 10.1109/CVPR.2013.178.

[28] Sheng Wang, Guohua Gou, Haigang Sui, Yufeng Zhou, Hao Zhang, and Jiajie Li. CDSFusion: Dense Semantic SLAM for Indoor Environment Using CPU Computing. *Remote Sensing*, 14(4), 2022. ISSN 2072-4292. doi: 10.3390/rs14040979. URL `https://www.mdpi.com/2072-4292/14/4/979`.

[29] Thomas Whelan, Stefan Leutenegger, Renato F. Salas-Moreno, Ben Glocker, and Andrew J. Davison. ElasticFusion: Dense SLAM Without A Pose Graph. In *Robotics: Science and Systems*, 2015.

[30] Shichao Yang and Sebastian Scherer. CubeSLAM: Monocular 3-D Object SLAM. *IEEE Transactions on Robotics*, 35(4):925–938, Aug 2019. ISSN 1941-0468. doi: 10.1109/TRO.2019.2909168.

[31] Hailin Yu, Youji Feng, Weicai Ye, Min Jiang, Hujun Bao, and Guofeng Zhang. Improving Feature-based Visual Localization by Geometry-Aided Matching. *ArXiv*, abs/2211.08712, 2022.

[32] Nicky Zimmerman, Tiziano Guadagnino, Xieyuanli Chen, Jens Behley, and Cyrill Stachniss. Long-Term Localization Using Semantic Cues in Floor Plan Maps. *IEEE Robotics and Automation Letters*, 8(1):176–183, Jan 2023. ISSN 2377-3766. doi: 10.1109/LRA.2022.3223556.