Grant agreement No: 101017008

Harmony

Assistive robots for healthcare

Enhancing Healthcare with Assistive Robotic Mobile Manipulation

(HARMONY) | H2020-ICT-2018-20| RIA

Start of the project: 01.01.2021
Duration: 42 months

Deliverable Number	D4.4
Deliverable Name	Change Detection for Long-term Mapping
WP Number	4
Lead Beneficiary	BONN
Dissemination Level	Public
Internal Reviewer	ETH
Due Date	30.04.2024
Date of Submission	
Version	1.0

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017008

Revision History

Version	Date	Author(s)	Comments
0.1	18-04-2024	Haofei Kuang	Initial draft
0.2	23-04-2024	Cyrill Stachniss, Jens Behley, Haofei Kuang	Revisions and corrections
1.0	30-04-2024	Haofei Kuang, Lionel Ott	Integrating comments and corrections by ETH

Table of Contents

Summary

In this deliverable, we describe our change detection approach within the Harmony scenarios. This deliverable discusses the development of object-based change detection and its integration into the mapping system for supporting long-term mapping. It supports the Harmony robot to operate effectively in complex and highly dynamic indoor environments and to maintain its world model.

In our work, we deploy the metric-semantic mapping algorithm described in Deliverable D4.2 as the representation of the world. The method exploits monocular RGB frames as input to recognize 3D objects through a monocular 3D object detection model and then integrates the objects' information to a given floor plan to build an object-based 2D map as the world representation.

Subsequently, our approach utilizes change detection at the object level between the current state of the world and a previously established metric-semantic map. We apply a data association method to match newly detected objects with the visible objects of the current camera view in the pre-built map. We categorize all objects into four change states depending on the conflict between the observation and the early world: new, removed, persistent, and unobserved. The change detection results allow for the continuous update of the map to support long-term mapping in dynamic environments. Our method uses only the RGB images reducing sensor setup requirements and integrates seamlessly with localization and SLAM systems in the Harmony project. The qualitative results show that the proposed change detection and map updating algorithm can adapt to daily changes in dynamic indoor scenarios.

Overall, the method proposed in this deliverable supports the long-term autonomous operation of the Harmony robot in highly dynamic hospital environments.

Acronyms

LiDAR	Light Detection and Ranging
ROS	Robot Operating System
SLAM	Simultaneous Localization and Mapping
SIMP	Semantic Indoor Mapping and Localization
PMT	Panoptic Multi-TSDFs

BONN	Rheinische Friedrich-Wilhelms-Universität Bonn
ETH	Eidgenössische Technische Hochschule Zürich

Introduction

Hospitals are highly dynamic and crowded environments consisting of a large number of people and moveable devices. Thus, a map needs to be regularly maintained and updated to provide real-time environmental information for the long-term operation of the robot. However, the mapping process is time-consuming and inefficient. To operate in such a complex environment efficiently, the Harmony robot needs to automatically detect the changes in the hospital to update the map to support other robotic tasks such as localization, navigation, and manipulation. It is a critical capability to allow the Harmony robot to reduce human intervention during long-term operations in the hospital. Therefore, the Harmony robot needs to keep track of the long-term changes in the environment during operation.

To address this problem, we developed an algorithm that allows robots to detect moved objects between the built map and recent observations. This involves tackling the challenge of detecting the changes between the current state of the world and a given environment model based on an object-centric map and the robot's perception. To achieve this goal, the project will extend the object-based mapping to object-centric change detection.

To obtain the environment model, we construct an object-based map through the metric-semantic mapping algorithm [1] developed by BONN. The algorithm is part of the Harmony SLAM system described in Deliverable D4.2 and it fuses 3D object detection results with an existing floor map as an object-centric map representation. It also deals with short-term changes by filtering dynamic objects. However, the method does not consider long-term changes in the environment which results in the environment model becoming outdated over time.

Following that, to detect the long-term changed object, we exploit the currently detected 3D objects as the current state of the world, and then we perform data association between the previously built objects and currently detected objects. Based on the data association results, we can detect conflicts between previously mapped objects and observed objects, i.e., the changes in the environment. Using these changes in object locations, we update the map to support the other tasks, such as localization.

Both parts of the mapping and change detection system have been built to utilize the Harmony sensor suite, i.e., our change detection framework leverages the forward-looking RGB-D sensor (only the RGB image is used here) and the wheel odometry.

In summary, by representing the world in an object-centric fashion, we can perform data association between the current world state and a previously built environment model to detect moved objects. Using such an object-based change detection potentially allows us to efficiently update the environment model to allow the Harmony robot to perform long-term mapping in a highly dynamic hospital environment.

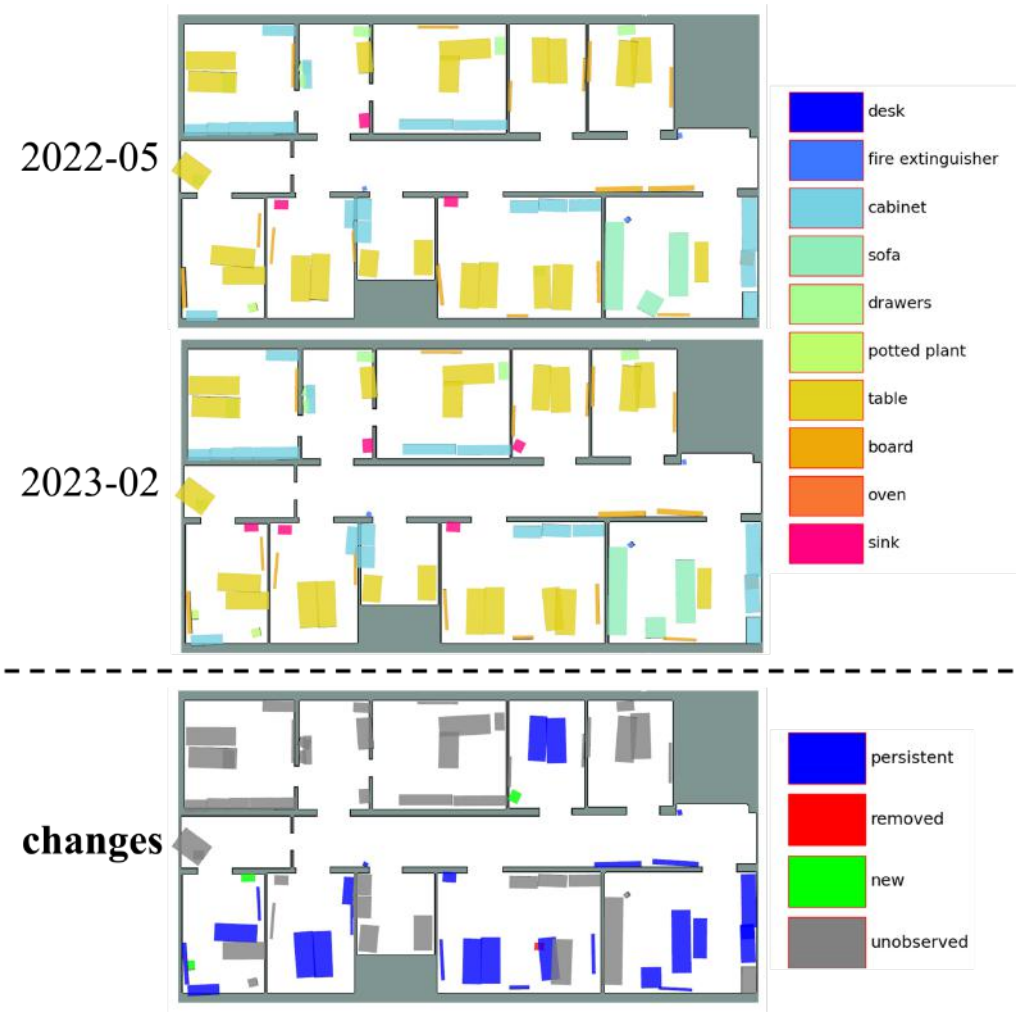


Figure 1: The change detection of two recorded datasets of our lab in BONN, which are more than a year apart. The upper part shows the semantic information provided by a 3D detection approach integrated into a floor plan. The bottom part shows the extracted changes differentiating between persistent, removed, added/new, and unobserved objects.

Change Detection and Map Update

In the Harmony scenarios, long-term precise mapping and localization are relevant capabilities of the autonomous system. To achieve these capabilities, the recognition and integration of changes to objects in the environment, enabled by object-based change detection and mapping, provides the basis for long-term operation in complex and highly dynamic indoor environments.

For this reason, we consider an approach that detects changed objects between the current state of the world and a given world model during long-term operation in the environment as shown in Figure 1. To align with the Harmony goals, we focus on data association between the previously built metric-semantic map and 3D object detections from monocular RGB frames.

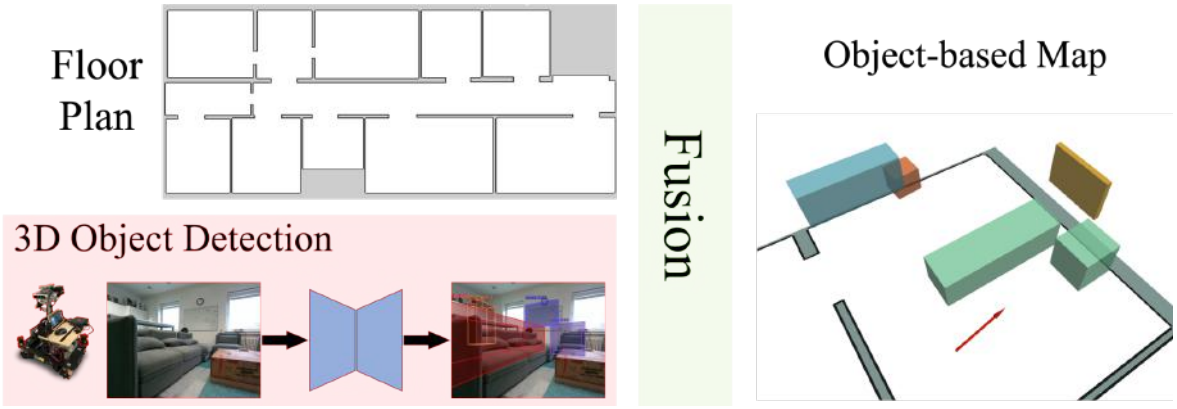


Figure 2: The mapping pipeline of SIMP. The object-based semantic map is constructed by performing 3D object detection and then fusing objects into a floor plan.

In the previous deliverable D4.2, we proposed a semantic indoor mapping and localization (SIMP) method [1] to construct an object-centric metric-semantic representation for the Harmony system. This mapping algorithm obtains the semantics by exploiting a 3D object detection model to predict 3D bounding boxes from monocular RGB images, which are then integrated into a floor plan, resulting in a metric-semantic representation of the environment, as shown in Figure 2. SIMP ideally combines the geometric and semantic information into a single map representation that can be easily used for downstream tasks, like localization (see D4.2). To support the highly precise localization performance, SIMP tries to reduce the effect of the dynamic objects in the environment. It filters the potentially short-term movable objects such as people or chairs by fine-tuning the 3D object detection model such that it focuses on unmovable objects such as tables or cabinets. According to that, SIMP can construct a clean and precise semantic map by handling the short-term changes in the environment, which supports downstream robotic tasks. Nonetheless, SIMP still ignores the fact that short-term unmovable objects can also change, e.g., objects can be moved, added, or entirely removed, over time.

Our approach presented here addresses and overcomes these limitations. Our approach detects long-term changes in terms of objects between the current state of the world and a pre-built metric-semantic map provided by SIMP. We treat the metric-semantic representation of SIMP as an object-centric map, i.e., each object will be treated as a submap and we perform change detection at the object level. Based on that, we still utilize only RGB frames to detect 3D objects of the current world state, which not only reduces the requirement of the sensor setup but also allows us to exploit the RGB camera of the Harmony sensor configuration. Then, we propose a data association method to match the

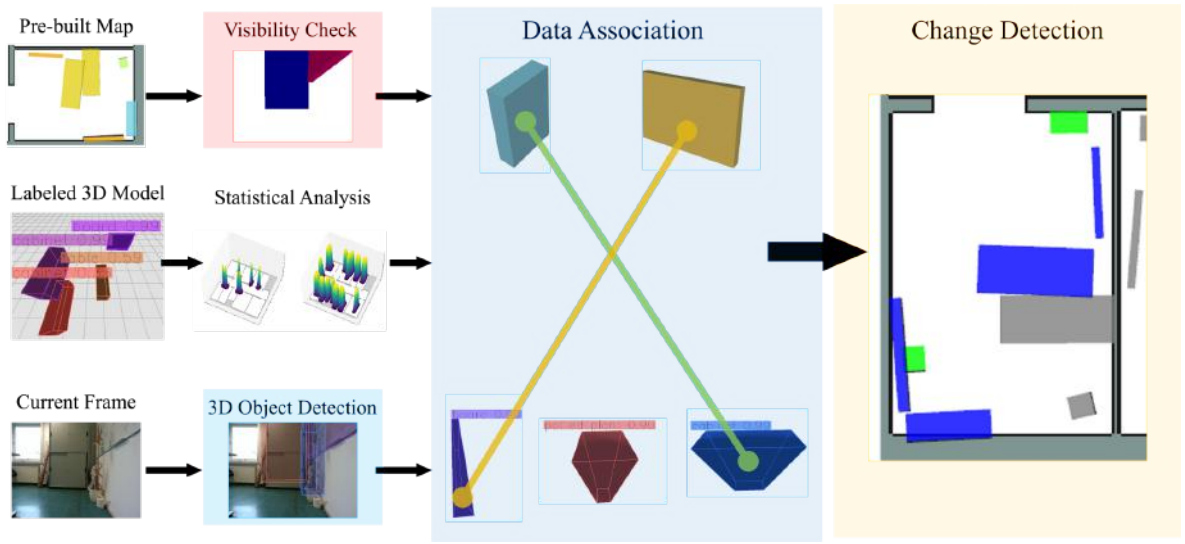


Figure 3: Overview of our change detection and long-term mapping system. There are three types of inputs for the data association: the visible map objects from the pre-built map on the current camera view, a pre-computed Gaussian distribution from the ground-truth 3D model, and the detected objects from the 3D object detection model. According to the data association results, we update each object's change state.

currently detected objects to the previously detected objects from an earlier point in time to obtain the match or conflict between the current observation and the previous map. The results of the data association lead to changes in the environments, e.g. the matched objects represent persistent, and unmatched objects represent new or removed. It allows us to perform the algorithm during the Harmony robot operating localization or navigation task to integrate the change detection results for efficiently updating the object-based map with the latest semantic information of the real world.

Figure 3 provides an overview of the proposed method. Regarding the perception and the object-based mapping module, we follow SIMP and exploit the monocular RGB frames and a floor plan as input, to construct the object-based semantic map, as shown in Figure 2. We also use a fine-tuned Cube R-CNN [2] network as a 3D object detector to predict the 3D bounding box of objects in the current RGB frame. Subsequently, the predicted 3D objects are fused and projected to the input floor plan to generate the final metric-semantic map. For more details about the object-based map representation refer to Deliverable D4.2.

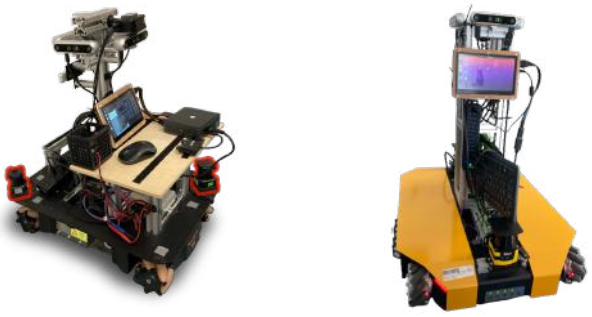
Regarding the long-term change detection, we extend SIMP to build the connection between the current new observation and the earlier built object-based map. In our system, the input consists of posed RGB images and the object-based map which was built by SIMP potentially a long time ago. Then, we first detect the 3D objects in the current RGB frames through Cube R-CNN and perform with these detections the data association between these newly detected objects and the existing objects in the previously built map. For convenience, we call the existing objects in the map: map objects. According to the data association results,

we can get the so-called change state of all objects, namely: new, removed, persistent, and unobserved. More specifically, (1) change state “new” means the current detected object did not match any map object, (2) change state “removed” means the map object should be observed in the current camera view, but it is not matched to any currently detected object, (3) change state “persistent” means the newly detected object matches to a map object, and (4) change state “unobserved” is an additional state for map objects that do not correspond to the current camera view, so their change state is currently unknown.

To achieve reliable data association, we build upon and use a previous Harmony contribution, called Panoptic Multi-TSDFs (PMT) [3] by ETH, and extend the method to fit our needs in the object-based map representation. PMT detects changes between two TSDF maps according to objects’ geometry and category. In our implementation, given a list of the detected objects and a list of the map objects, we first project the existing map objects to the current camera view through ray tracing. A map object is set to active if it is visible in the current view, i.e., the rendered 2D bound box of the object in the image plane. To compute the data association cost between two objects, we first transfer the 3D bounding box of a detected object to a map object according to its relative center position and direction. Then, we compute the intersection over union (IoU) between these two bounding boxes. We also perform the statistical analysis, which is described in SIMP, to compute the category-aware distribution of each object and then calculate the probability distance between them. Specifically, each object has a pre-computed Gaussian distribution according to its category. The similarity between a map object and a detected object is the probability in the map object’s distribution queried by the center of the detected object. It provides a way to determine whether the two objects belong to the same instance. More details of the statistical analysis are described in our paper [1]. After data association, we can detect the conflict between the current observation and the pre-built map, then we update the change state for each object and also the object-based map. Compared with PMT, the proposed method does not rely on the depth information from an RGB-D sensor to reduce the requirements of the sensor setup.

Besides, our metric-semantic map representation is much more efficient and provides lower memory consumption than the TSDF-based volumetric map, and is also easier to integrate with the Monte-Carlo Localization [4] framework and Harmony SLAM system described in D4.1 and D4.2, respectively.

We evaluated our approach qualitatively on data collected with our platform Youbot and Ding-O in our lab in Bonn, Germany. These platforms replicate the Harmony robot sensor setup and actuation model. Both platforms are equipped with an Intel NUC10i7FNK and four Intel RealSense D455 RGB-D cameras (front, back, left, and right, but only the front camera is used here), as shown in Figure 4.



(a) KUKA YouBot (b) Clearpath Dingo-O

Figure 4: The two mobile platforms in BONN. Both sensor setup follows the Harmony sensor setup described in Deliverable D3.2.

To record naturally occurring changes in the environment, we record data in our lab over a long period of time, ranging from several months to a year. The qualitative results of the change detection are shown in Figure 5 and Figure 6. In Figure 5, the robot can detect the new objects (green bounding boxes) that are present in the current observation but not in the pre-build map. These objects might not exist before or have not been observed before because of occlusion. In Figure 6, a cabinet is detected as a removed object (red bounding box) because of the occlusion of the current observation. Besides, most objects in the environments are detected as persistent (blue bounding boxes), they might have a slightly different position but we still recognize them as persistent as long as they are consistent at the semantic level.

Our change detection approaches run about 5 Hz on our platform, and the 3D object detection runs about 9 FPS. We asynchronously execute change detection and 3D object detection to achieve real-time, e.g. we run 3D object detection for each frame and perform change detection when the robot moves a predetermined distance.

Figure 7 shows the procedure of the change detection and long-term mapping in our lab in BONN from January 2022 to February 2023. Additionally, we further showcase the system in action in a live demo and provide a video showing the change detection on the data collected in Bonn using our platform.

In summary, we implemented and demonstrated an algorithm that allows us to detect changes in the environment between the current state of the world and a given world model only using the monocular RGB frames. Results on our lab data show promising performance in indoor environments that change daily with human activities.

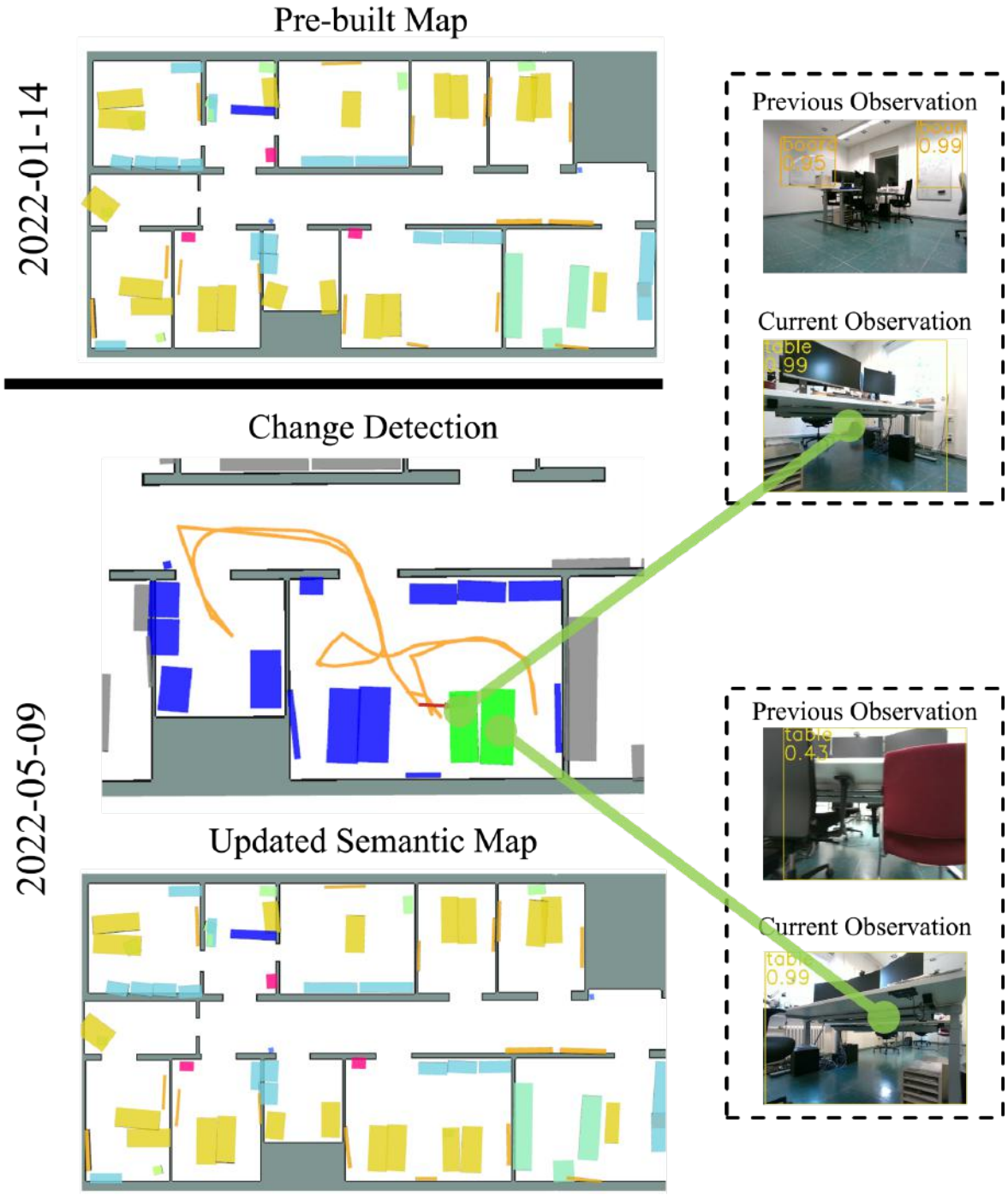


Figure 5: Detecting new objects in our lab environment. In the change state map (middle), the orange line is the robot's trajectory, the blue bounding box represents the persistent objects, the green bounding boxes are the newly observed objects, and the grey bounding boxes are the unobserved objects. We show the difference between previous and current observations.

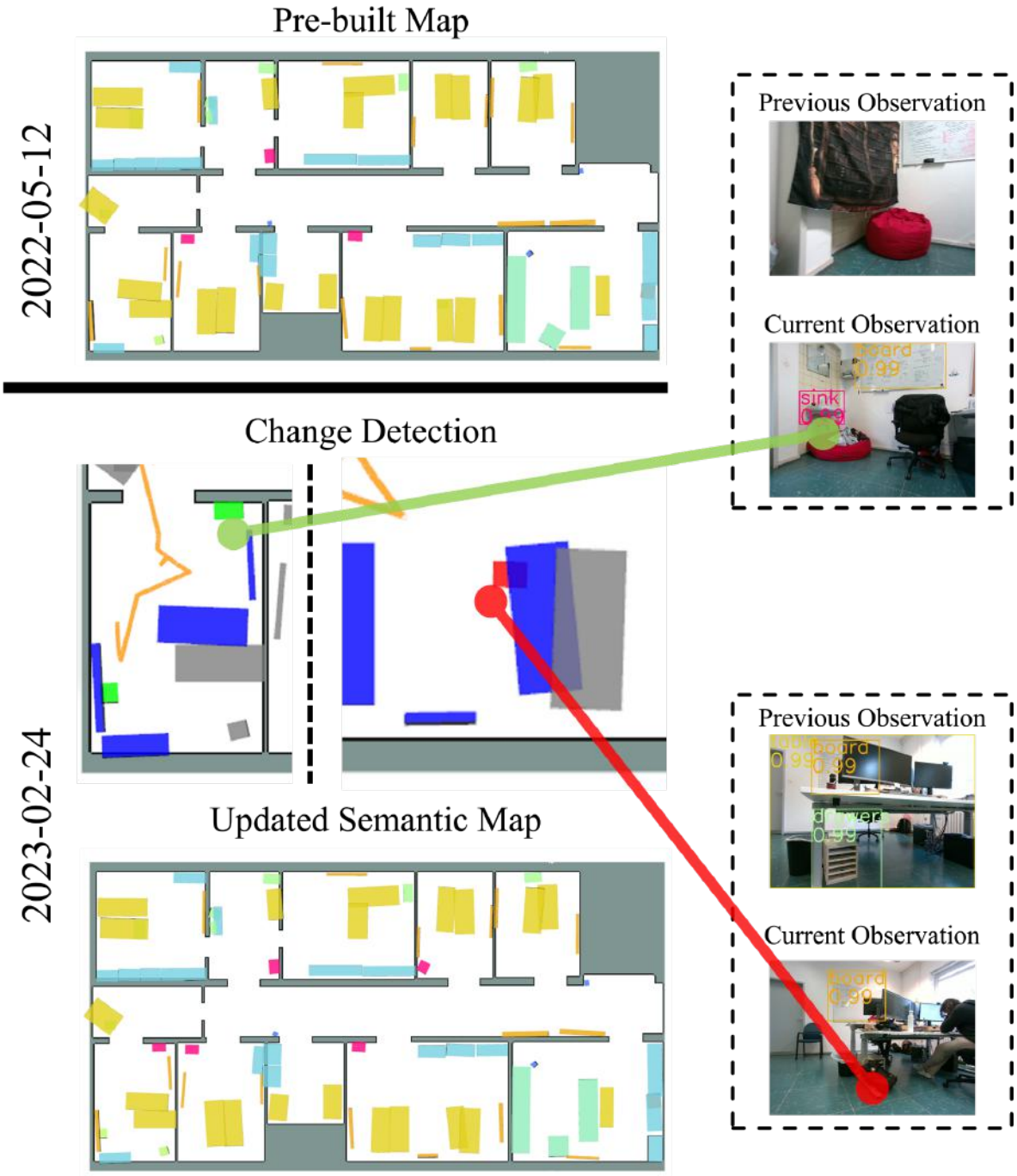


Figure 6: Change detection across almost a year in our lab environment. In the change state map (middle), the orange line is the robot's trajectory, the blue bounding box represents the persistent objects, the green bounding boxes are the newly observed objects, the red bounding box is the removed object, and the grey bounding boxes are the unobserved objects. We show the difference between previous and current observations.

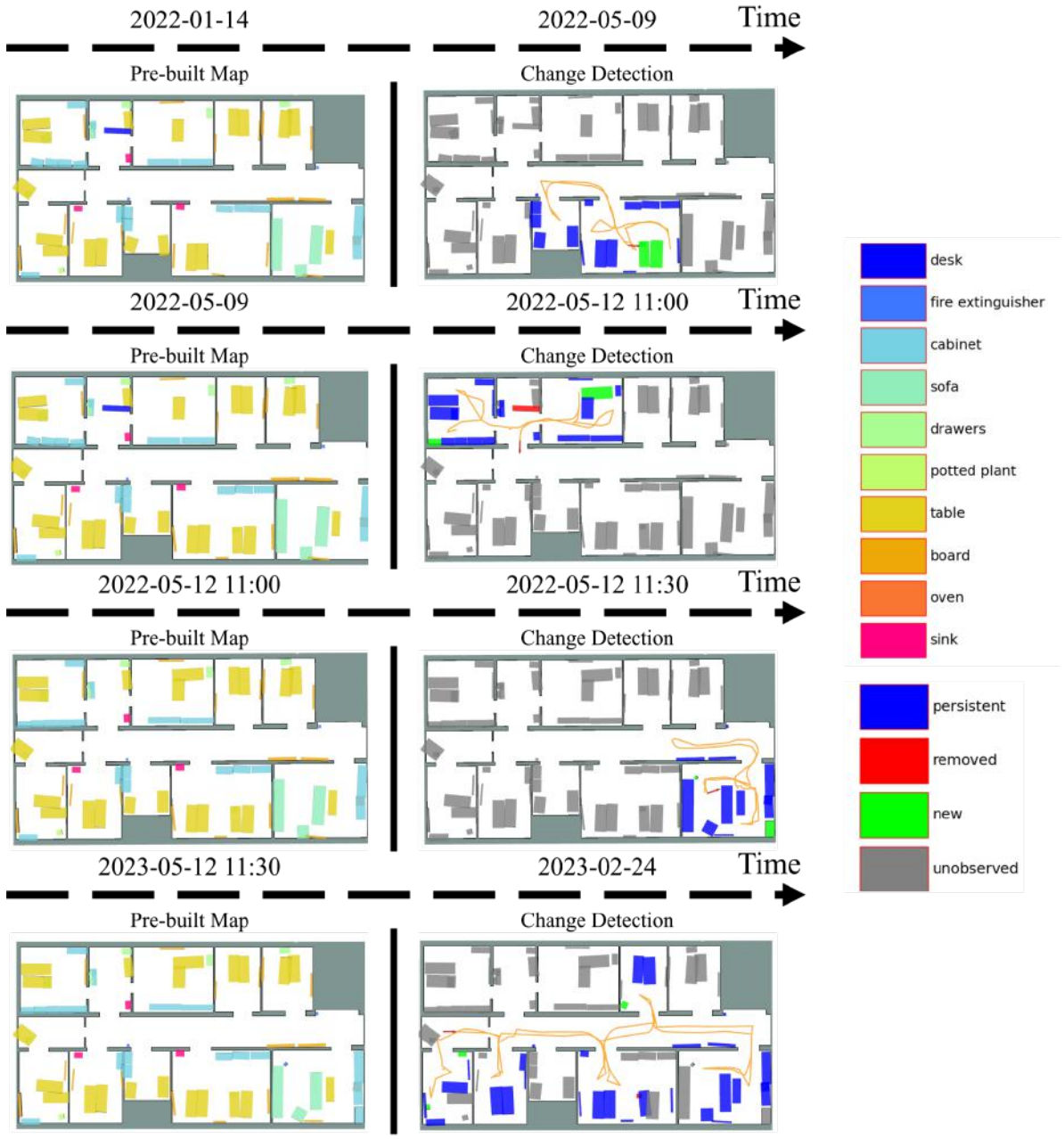


Figure 7: The change detection and long-term mapping from January 2022 to February 2023 at the lab in BONN. In the office environment, the green bounding boxes mean the new objects, the red bounding boxes are the removed objects, the blue bounding boxes are the persistent objects, and the grey bounding boxes mean that the map objects are not observed at this time. The left part of the figure shows the progressively updated semantic map of the current environment.

Software Installation

The change detection method is intended for Ubuntu 20.04 with ROS 1 Noetic, and Ubuntu 22.04 with ROS 2 Humble, and was tested in these settings. Detailed instructions outlining the installation, additional requirements, and execution instructions will be shared with the Harmony partners in the Harmony repository (we will provide the code and instructions in May 2024). Trained models for 3D object detection, as well as README files with instructions for ROS development, can also be found in the same repository.

Conclusion

In conclusion, our presented change detection and map update algorithm, which is based on the object-based map representation effectively addresses the challenges of long-term mapping in dynamic indoor environments. It is the result of jointly built software at BONN and ETH, building consistently upon prior Harmony developments. By leveraging monocular 3D object detection and developing a data association method with geometric and semantic information, we have successfully detected the long-term changes in the environment from the current state of the world and an early pre-built world model. It allows our system to handle the natural daily changes accurately and efficiently, thereby supporting downstream tasks like localization and navigation in complex hospital scenarios. Our extensive evaluations conducted over a year of natural changes at Bonn confirm our approach can effectively detect the daily changes due to human activities.

References

[1] N. Zimmerman, M. Sodano, E. Marks, J. Behley, and C. Stachniss. Constructing Metric-Semantic Maps using Floor Plan Priors for Long-Term Indoor Localization. Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS), 2023.

[2] G. Brazil, A. Kumar, J. Straub, N. Ravi, J. Johnson, G. Gkioxari. Omni3D: A Large Benchmark and Model for 3D Object Detection in the Wild. Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023.

[3] L. Schmid, J. Delmerico, J. L. Schönberger, J. Nieto, M. Pollefeys, R. Siegwart, C. Cadena. Panoptic Multi-TSDFs: a Flexible Representation for Online Multi-resolution Volumetric Mapping and Long-term Dynamic Scene Consistency. Proc. of the IEEE Intl. Conf. on Robotics and Automation (ICRA), 2022.

[4] F. Dellaert, D. Fox, W. Burgard, and S. Thrun. Monte Carlo Localization for Mobile Robots. In Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA), 1999.