

# A Coherent Semantic Mapping System Based on Parametric Environment Abstraction and 3D Object Localization

Ziyuan Liu      Wei Wang      Dong Chen      Georg von Wichert

**Abstract**—An important prerequisite for intelligent robots to effectively perform daily tasks in indoor environments is an advanced environment model. Not only should this model contain metric level information about the geometry of the perceived environment, it should also provide abstract level information, such as topology and objects, so as to benefit high level robotic applications. In this paper, we propose a mapping system which generates a coherent semantic map for indoor environments while taking laser range data and RGB-D images as sensor input. We propose to realize this system by combining parametric environment abstraction, i.e. using a parametric model to approximate the geometry of the perceived environment, with 3D object localization. Experiments using real world data show promising results and thus confirm the usefulness of our system.

## I. INTRODUCTION

To communicate with humans, a robot assistant needs to understand its operating environments not only in the metric sense but also on the semantic level. For instance, we want a robot assistant to fetch a tool for us that is located somewhere in another room. We humans will not (are also not able to) tell the robot the exact 3D coordinates of the tool or those of the room accommodating the tool. Instead, we will probably say something like “go to the *lab* and get me the *hammer* on the *table*”. In this case, *lab* and *table* are semantic terms describing the operating environment, and *hammer* is a semantic tag of the objects that are already known to the robot. Given this semantic information, the robot can plan a path to the *lab* and search for the *hammer* on the tables of the *lab*. On the other hand, the robot also needs the metric information of the environment to localize itself, to navigate to the goal and finally to look for the desired object. Therefore, it is obvious that a coherent semantic map which provides both semantic level understanding and metric representation of the operating environment is very important for intelligent robots to successfully and efficiently perform daily tasks.

In this paper, we extend our previous work [12] with 3D object localization [24]. The new system as a whole demonstrates a coherent semantic mapping approach for indoor environments that combines parametric environment abstraction and rich object information. The remainder of this

paper is constructed as follows: in section II we review the related work on semantic robot mapping. In section III we provide an overview of our system. In section IV, section V and section VI we explain the important components of our system. In section VII we show experimental results. Finally, we conclude in section VIII.

## II. RELATED WORK

A big group of literature focused on semantic labelling processes which divide the perceived environment into several semantically annotated regions (like “kitchen” or “office”) or into parts with different semantic meanings (like “walls”, “floors”, “ceilings”). In [16] a semantic labelling process is proposed for domestic environments which used the information about doors as an important feature. Pronobis and Jensfelt [18] integrated multi-modal sensory information, human intervention and some common-sense knowledge to classify places with semantic types. In [14], a logic-based constraint network describing the relations between different parts is used for labelling indoor environments. Combining range data and omni-directional images, buildings and natural objects in an outdoor scenario are labelled by semantic tags in [17]. Based on range images, Zhu et. al. [26] implemented a semantic labelling system on a vehicle to classify urban scenes. Other examples of this kind can be found in [8], [11], [1] and [21]. In general, the output of these labelling processes is a labelled data set, such as labelled grid maps in 2D cases and labelled point clouds, plane segments in 3D cases. Although they can semantically explain the perceived environments, they do not provide a parametric model which compactly describes the geometry of the environments.

Another group of related work concentrated on generating topological maps of the environments traversed by a robot. Blanco et. al. [4] proposed a hybrid metric-topological approach based on Bayesian inference. In addition to traditional metric maps, their approach was able to reconstruct the path of the robot as a topological map. [22] demonstrated a topological SLAM process which used the saturated generalized Voronoi graph (S-GVG) to represent the topology of an office-like environment. Based on particle filtering, Werner et. al. [25] proposed to build topological maps for an indoor environment by disambiguating places which appear indistinguishable using neighbourhood information extracted from a sequence of observations. Other examples in this group can be found in [3], [7] and [2]. Approaches in this category were mainly interested in building globally consistent metric maps while providing topology along the

Z. Liu and D. Chen are with the Institute of Automatic Control Engineering, Technische Universität München, D-80290, Munich, Germany. ziyuan.liu@tum.de, chendong@mytum.de

W. Wang is with the Faculty of Informatics, Technische Universität München, D-80290, Munich, Germany. wei.wang@in.tum.de

G. von Wichert is with Siemens AG, Corporate Technology, Munich, Germany and Institute for Advanced Study, Technische Universität München, Munich, Germany georg.wichert@siemens.com

robot path as a by-product. In this sense, they represent the early tries toward semantic mapping.

A third category of literature consists of object-based semantic mapping systems which use objects as basic representation units of the perceived environment. Such systems usually adopt point cloud processing and image processing techniques to model or detect objects. Object features like appearance, shape and 3D locations are often used to represent the objects. Using 3D point cloud data, Rusu et. al. [20] proposed a hybrid semantic object mapping system for household environments (mainly kitchens). Objects modelled in this work are those which perform utilitarian functions in the kitchen, such as kitchen appliances, cupboards, tables, and drawers. More examples of object-based semantic mapping can be found in [19], [15], [6] and [13]. Approaches of this kind mainly focused on describing the perceived environments (usually small environments, e.g. kitchens) on object level (using object positions and relations) and did not interest in modelling the geometry of the environments.

### III. SYSTEM OVERVIEW

Our system takes laser range data and RGB-D images of a Kinect camera as input. As output, it provides a coherent semantic map of the perceived environment which combines parametric environment abstraction and 3D object localization. Fig. 1 shows an overview of our system. The laser range data is firstly processed by a grid mapping algorithm to generate an occupancy grid map of the environment and to provide a coherent global coordinate system. In our work, we have used the GMapping [9] algorithm for this purpose. The resulting grid map is then used as input for the process of parametric environment abstraction [12] which uses rectangular space units to approximate the geometry and the topology of the perceived environment. Within each space unit, unknown areas of the grid map are detected based on connected-components analysis [5]. Such areas are considered as obstacles which can not be traversed by robots. On the other hand, 3D objects that are captured by RGB-D images of the Kinect camera are recognized and localized by the algorithm in [24]. Finally, a coherent semantic map that captures the geometrical, topological and object information of the operating environment is generated by incorporating the 3D object information into the parametric environment model.

In the coherent semantic map shown in Fig. 1, ellipse nodes represent space units, each of which has a rectangular shape. Black solid edges indicate that two space units are connected by one or more doors. Black dashed edges imply that two space units are only adjacent (neighbour) yet not connected. Blue rectangles show the detected 3D objects. Blue edges show the belongingness of these objects, i.e. which space units these objects belong to. Red rectangles and red edges indicate the obstacles detected in the space units. These obstacles are drawn as green color blobs in the parametric model.

Given such a map, a robot not only knows the geometry and topology of the environment, it also has a prior belief

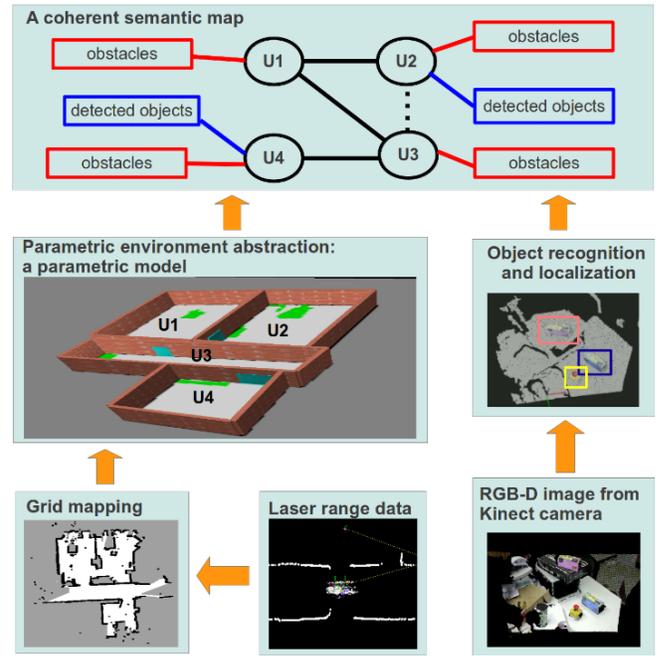


Fig. 1. System overview. In the coherent semantic map, ellipse nodes represent space units. Black solid edges indicate connectivity by doors. Black dashed edges imply adjacency (not connected by doors). Blue edges indicate the belongingness of the detected 3D objects. Red rectangles and red edges show the obstacles detected in the space units which are illustrated by green color blobs in the parametric model.

on the pose of objects and their belongingness. Furthermore, this map provides a robot semantic understanding of the underlying environment, which can benefit high level robotic applications. For instance, if a robot needs to plan a path from position  $A$  that is located in space unit  $U1$  to position  $B$  in unit  $U4$ , instead of a brute-force path planning  $A \rightarrow B$ , which happens directly on the coordinate level, the robot can plan a path based on the topology given by the coherent semantic map. As shown in Fig. 1, the actual path planning can be divided into a few parallel steps on the semantic level:  $A \rightarrow \text{door}(U1, U3)$ ,  $\text{door}(U1, U3) \rightarrow \text{door}(U3, U4)$  and  $\text{door}(U3, U4) \rightarrow B$ . Since the positions  $\text{door}(U1, U3)$  and  $\text{door}(U3, U4)$  are also given by the semantic map, the above path planning steps can be easily done using standard methods like the A\* algorithm [10].

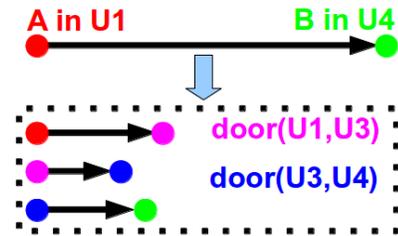


Fig. 2. Path planning on the semantic level. Instead of a brute-force planning from  $A$  to  $B$ , the robot can plan a path based on the topology given by the semantic map in a few parallel steps.

#### IV. PARAMETRIC ENVIRONMENT ABSTRACTION

The fundamental idea is to use a parametric model to approximate the geometry of indoor environments, so that we can use an abstract model to represent the perceived environment. This parametric model should have a simple geometrical form which is easy to use and change. On the other hand, it should have the power to accurately represent environments. Following the knowledge that space units of indoor environments, such as rooms and corridors, have mostly a rectangular shape, we decide to use rectangles as basic elements of our parametric model to represent these units. We call each instance of our parametric model a semantic world  $W$ . Let  $M$  be the grid map that captures the perceived environment. Our aim is then to find the best semantic world  $W^*$  which fulfils the following condition:

$$W^* = \operatorname{argmax}_{W \in \Psi} p(W|M), \quad (1)$$

with  $\Psi$  being the solution space.

The *posterior* probability  $p(W|M)$  is computed as

$$p(W|M) \propto p(M|W) \cdot p(W), \quad (2)$$

where  $p(M|W)$  is the *likelihood* measuring how well the map  $M$  matches with the semantic world  $W$ .  $p(W)$  is the *prior* probability describing what kind of semantic worlds are possible at all.

In order to solve the equation (1), we formulate the whole problem as a Markov chain. Each state of the Markov chain represents a semantic world. Starting from an initial guess of the semantic world, we adopt a data driven MCMC [23] sampling process to change the Markov chain state, so that the underlying semantic world better matches the input map. For this purpose, we use several reversible kernels (shown in Fig. 3) to change the geometry of the space units. An example of the whole parametric abstraction procedure is shown in Fig. 4. More details on parametric environment abstraction can be found in [12].

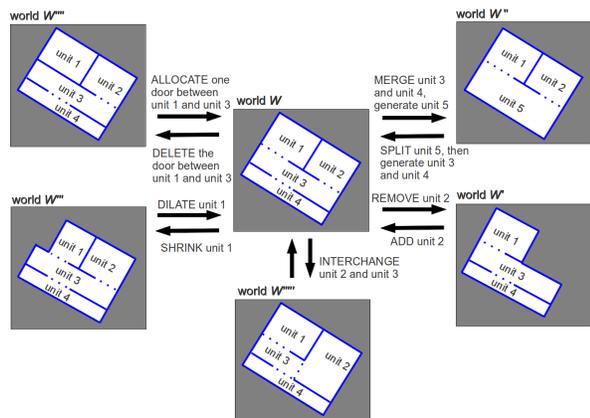


Fig. 3. Reversible MCMC kernels: ADD/REMOVE, SPLIT/MERGE, SHRINK/DILATE, ALLOCATE/DELETE and INTERCHANGE. The world  $W$  can transit to  $W'$ ,  $W''$ ,  $W'''$ ,  $W''''$  and  $W''''''$  by applying the kernel REMOVE, MERGE, SHRINK, DELETE and INTERCHANGE, respectively. By contrast, the world  $W'$ ,  $W''$ ,  $W'''$ ,  $W''''$  and  $W''''''$  can also transit back to  $W$  using corresponding reverse kernel.

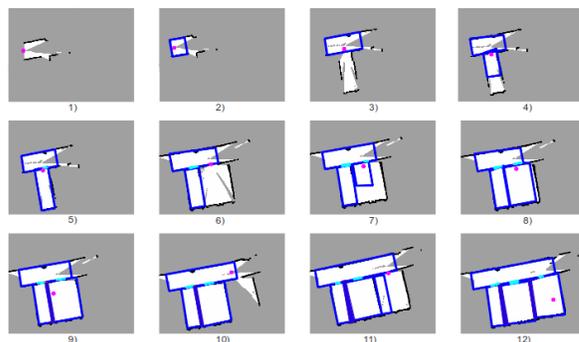


Fig. 4. An example of parametric environment abstraction. Using the MCMC kernels shown in Fig. 3, the semantic world is adapted to better match the input map. Here the semantic world (blue=walls,cyan=doors,magenta= current robot pose) is directly plotted on the input map (black=occupied,gray=unknown,white=free).

#### V. 2D OBSTACLE DETECTION

Since there exist many things in real world which are difficult for a robot to recognize, such as furniture, it is reasonable to model these things as general obstacles, so that the robot knows that it can not traverse them and should keep certain distance to avoid collision. Given the parametric environment model described in section IV, we propose to detect such obstacles using connected-components analysis [5]. Within each space unit of a semantic world, we detect connected unknown cells of the input map as potential obstacles. If the area of such potential obstacles are bigger than certain threshold, we consider them as real obstacles. An example of obstacle detection is depicted in Fig. 5.

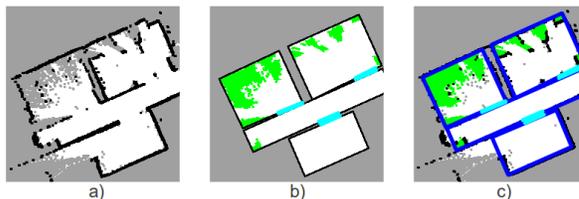


Fig. 5. 2D obstacle detection. **a)** The input grid map (black=occupied, gray=unknown, white=free). **b)** The corresponding semantic world (black=walls, gray=unknown, white=free, cyan=doors, green=obstacles). **c)** A direct comparison between the input map and the semantic world. The walls of the semantic world is drawn in blue. Other colors mean the same as in a) and b).

#### VI. 3D OBJECT LOCALIZATION

For object localization, we use an object recognition and pose estimation framework which is based on a global object descriptor called *Viewpoint oriented Color-Shape Histogram* (VCSH) [24]. This framework uses both object's color and shape features to recognize textured and textureless objects. In addition, it is able to retrieve accurate 6D poses of recognized objects.

The complete 3D object model  $\Omega$  is built by registering all the object's RGB-D data of different poses into a single coordinate frame. Using the object model's centroid as origin, we generate a sphere with a certain radius. On the surface of

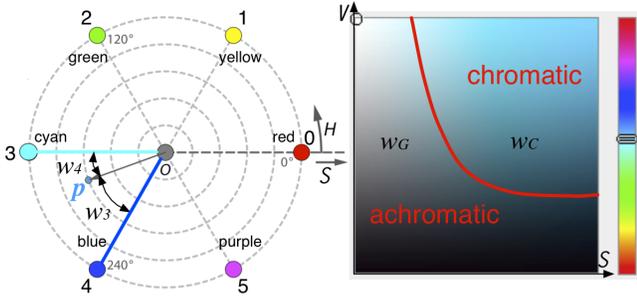


Fig. 6. Left: smoothed color range and estimation of the contributions for neighbor regions in  $HS$  space. Right: illustration of the chromatic and achromatic areas in  $SV$  space.

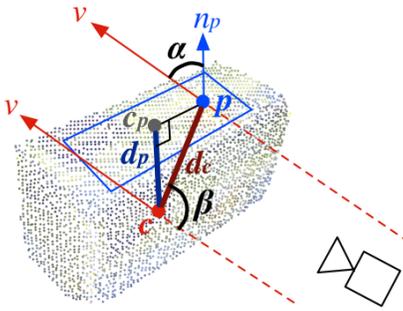


Fig. 7. Shape features of point  $p$ .  $c$  is the centroid of object path data.  $n_p$  is the normal of  $p$ .  $v$  is the synthetic viewpoint direction.  $c_p$  is  $c$ 's projection point on the tangent plane of  $p$  (blue rectangle frame).  $d_c$  and  $d_p$  are the distances from  $c$  to  $p$  and from  $c$  to  $c_p$ .  $\alpha$  is the angle between  $v$  and  $n_p$ , and  $\beta$  is the angle between  $v$  and the line segment  $cp$ .

this sphere, a big amount of viewpoints are homogeneously generated with their direction pointing to the origin of this sphere. Using each of these generated viewpoints  $VP$ , an object patch datum  $M$  which represents the object label and the corresponding viewpoint pose, is generated accordingly using a ray-casting like method. Subsequently, we generate a VCSH for each object patch data  $M$ . Finally, an object is represented by the generated VCSH set.

#### A. VCSH building based on color and shape features

a) *Color features*: The  $HSV$  value of each point in object path data is used to estimate color contributions both in chromatic and achromatic area. As shown in Fig. 6, the chromatic area is considered as the true color space, and achromatic area represents the gray scale space in the whole  $SV$  space. Eight histogram regions  $RE$  with index  $u = \{0 \dots 7\}$  are defined for building VCSH. Six of them are used for chromatic area, and the other two are for achromatic area. According to the  $HSV$  value, each point  $p$  is ranged into two adjacent histogram regions  $\langle RE_u, RE_{u+1} \rangle$  in chromatic area, and one region  $RE_6$  or  $RE_7$  in achromatic area, with respective contributions being  $\langle w_u, w_{u+1}, w_6 | w_7 \rangle$ . An example is shown in Fig. 8.

b) *Shape features*: To extract shape features of point  $p$  in the object patch data, we project the centroid  $c$  of the patch data as a point  $c_p$  to the tangent plane of  $p$ . As shown

in Fig. 7, we define the shape feature of point  $p$  as  $f := \langle d_p, d_c, \alpha, \beta \rangle$ , with  $\langle d_p, d_c \rangle$  being two distance components and  $\langle \alpha, \beta \rangle$  being two angle components. For a single object model which contains  $J$  object patch data, the  $j$ th shape feature set  $F_j = \{f_0 \dots f_m\}$  with  $m$  points represents the object's shape from the  $j$ th viewpoint  $VP_j$ .

c) *VCSH*: An example of generating VCSH using color and shape features is illustrated in Fig. 8. We use constant histogram bins for each shape feature in histogram regions  $RE$ . The shape features of each point  $p$  are separately indexed into a certain bin as  $\langle INX_{d_p}, INX_{d_c}, INX_{\alpha}, INX_{\beta} \rangle$ . The color contributions  $\langle w_u, w_{u+1}, w_6 | w_7 \rangle$  of each point in an object patch data are added to shape feature's indexes of the corresponding three histogram regions  $\langle RE_u, RE_{u+1}, RE_6 | RE_7 \rangle$ . Finally, these extracted color and shape features are correlated as a VCSH to represent the corresponding object patch data. A complete 3D object model is represented by a big VCSH set.

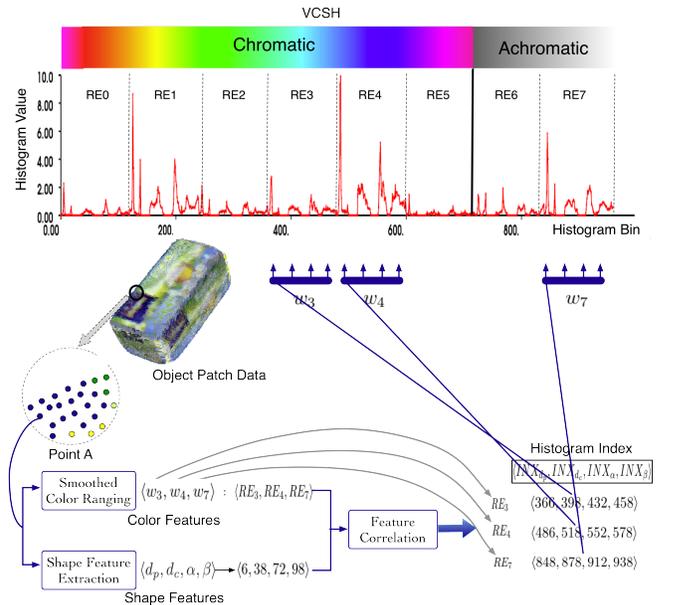


Fig. 8. Generation of a VCSH using color and shape features of an object patch data. Each object patch data is represented as a VCSH and stored into the database. A complete object model is represented by a big VCSH set.

#### B. Object Recognition and Pose Estimation

To recognize 3D objects and to estimate their correct 6D poses in real scenes, our framework firstly segments the background in the input RGB-D images. The foreground data are then grouped into different object clusters  $C_k$ ,  $k \in [1, N]$ , according to Euclidean distance. In this case, each cluster  $C_i$  is essentially a patch data of a perceived object and can be assigned a VCSH. Based on kd-tree, we use chi-squared distance  $\chi^2$  between each cluster's VCSH value  $H(C_k)$  and each VCSH value  $H_{ij}$  in the trained database to search for the best matching. The best matched object identification and the relative viewpoint pose  $\langle L, \hat{P} \rangle$  are extracted as:

$$\langle L, \hat{P} \rangle = \arg \min_{\langle L, P \rangle_{ij}} \chi^2(H(C_k), H_{ij}), \quad (3)$$

where  $i$  is the object model index, and  $j$  is the synthetic viewpoint index.

Subsequently, an iterative closest point (ICP) method and a pose verification method are used to get an accurate pose estimation of the object. The object's pose is then transformed into the semantic map.

## VII. EXPERIMENTS AND DISCUSSIONS

Fig. 9 shows the mobile robot platform that we have used to perceive indoor environments. This robot is equipped with a Kinect camera, a stereo camera system and three laser scanners. For our experiments we mainly used the Kinect camera and the two laser scanners which are situated at the front and back side of the robot respectively.



Fig. 9. Our mobile robot platform performing daily tasks, such as localization, objects detection, grasping and so on. Equipped sensors: three laser scanners, a Kinect camera and a stereo camera system.

The performance of our system applied to an indoor office environment which is comprised of six space units (five office rooms and one corridor) is illustrated in Fig. 10. As shown in Fig. 10-a, the resulting coherent semantic map correctly interprets the perceived environment with space units  $U_1, U_2 \dots U_6$  and a corresponding topology (connectivity by doors and adjacence). Using our parametric model, the grid map of the perceived environment is divided into six rectangular regions (space units). Based on these regions, we can distinguish between obstacles located in the units and the background (unexplored area) of the environment, both of which are comprised of gray cells in this case. 2D obstacles detected in the space units are plotted in green in the corresponding grid map. As we can see in Fig. 10-b, these obstacles essentially represent the furniture of the perceived environment, such as tables and cabinets. By setting a proper height to walls (blue) and doors (green), we show the 3D parametric model along with the detected 3D objects in Fig. 10-c. Here the table planes, on which the 3D objects are detected, are shown in cyan. The current robot pose (represented by a cluster of coordinate systems), the current laser scan (shown by red points) and the current color point cloud are highlighted by the dashed yellow rectangle.

Fig. 10-d depicts the details of object recognition and localization. In space units  $U_1, U_2$  and  $U_5$ , several 3D objects are recognized and localized regarding their 6D pose ( $x, y, z, \text{roll}, \text{pitch}, \text{yaw}$ ). Our object recognition and pose

estimation framework achieves a success rate of 92%. The estimation error of the 6D object pose is under 24 mm in translation and 1.6 degree in rotation. For a single object, our framework takes less than 1 second for recognition and pose estimation. Moreover, our VCSH is efficient and stable enough under varying illumination conditions in common environments.

By cell-wise checking of our parametric model and the input grid map, we measured an accuracy of 94.1% for geometry approximation. The mismatch of 5.9% is mainly due to some not-fully-explored areas of the input map, which are too small to be recognized as a space unit. Such areas indicate the existence of space units that were only partially explored during the experiment.

## VIII. CONCLUSION

In this paper we proposed a semantic mapping approach for indoor environments using laser range data and RGB-D images. Our mapping system provides a coherent semantic map of the perceived environment which not only well explains the environment on the abstract level but also accurately approximates the environment geometry using a parametric model. Moreover, 2D obstacles and 3D objects are modelled as abstract terms in our semantic map in a consistent way. By providing such coherent semantic maps, we aim to provide robots necessary information which is required for successful execution of daily tasks.

## ACKNOWLEDGEMENTS

This work is accomplished with the support of the Technische Universität München - Institute for Advanced Study, funded by the German Excellence Initiative.

The work of one of the authors (Georg von Wichert) was partially made possible by funding from the ARTEMIS Joint Undertaking as part of the project R3-COP and from the German Federal Ministry of Education and Research (BMBF) under grant no. 01IS10004E.

## REFERENCES

- [1] S.Y. An, L.K. Lee, and S.Y. Oh. Fast incremental 3d plane extraction from a collection of 2d line segments for 3d mapping. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012.
- [2] Adrien Angeli, Stéphane Doncieux, J-A Meyer, and David Filliat. Incremental vision-based topological slam. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1031–1036. IEEE, 2008.
- [3] Adrien Angeli, Stéphane Doncieux, J-A Meyer, and David Filliat. Visual topological slam and global localization. In *International Conference on Robotics and Automation*, pages 4300–4305. IEEE, 2009.
- [4] J-L Blanco, J-A Fernández-Madrugal, and Javier Gonzalez. A new approach for large-scale localization and mapping: Hybrid metric-topological slam. In *IEEE International Conference on Robotics and Automation*, pages 2061–2067. IEEE, 2007.
- [5] Fu Chang, Chun jen Chen, and Chi jen Lu. A linear-time component-labeling algorithm using contour tracing technique. *Computer Vision and Image Understanding*, 93:206–220, 2004.
- [6] F.R. Corrêa and J. Okamoto. Semantic mapping with image segmentation using conditional random fields. In *International Conference on Advanced Robotics*, pages 1–6. IEEE, 2009.

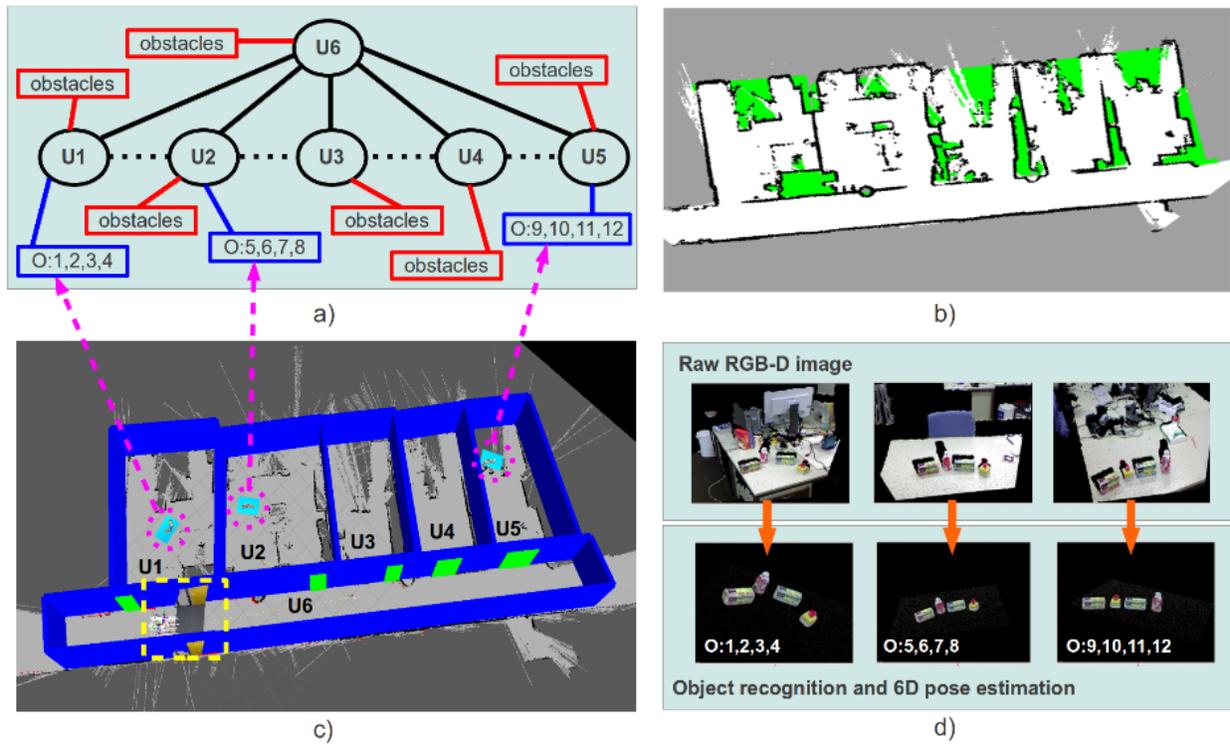


Fig. 10. Performance of our system using real world data. **a)** The coherent semantic map. Black ellipses indicate space units, each of which has a rectangular geometry. Solid black edges mean that two space units are connected by one or more doors. Dashed edges imply that two space units are adjacent but not connected by doors. Blue rectangles show the detected 3D objects. Blue edges show the belongingness of these objects. Red rectangles and red edges indicate obstacles detected in a unit. **b)** The resulting grid map of the perceived environment. The detected obstacles are drawn in green. **c)** By setting a proper height to walls and doors, we plot the 3D parametric model directly onto the corresponding grid map (blue=walls, green=doors, cyan=detected tables with 3D objects). The current robot pose (represented by a cluster of coordinate systems), the current laser scan (shown by red points) and the current color point cloud are highlighted by the dashed yellow rectangle. **d)** Details on 3D object localization.

[7] F. Ferreira, I. Amorim, R. Rocha, and J. Dias. T-slam: Registering topological and geometric maps for robot localization in large environments. In *International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 392–398. IEEE, 2008.

[8] N. Goerke and S. Braun. Building semantic annotated maps by mobile robots. In *Proceedings of the Conference Towards Autonomous Robotic Systems*, 2009.

[9] G. Grisetti, C. Stachniss, and W. Burgard. Improved techniques for grid mapping with rao-blackwellized particle filters. *IEEE Transactions on Robotics*, 23(1):34–46, 2007.

[10] Peter E Hart, Nils J Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. *Systems Science and Cybernetics, IEEE Transactions on*, 4(2):100–107, 1968.

[11] A.K. Krishnan and K.M. Krishna. A visual exploration algorithm using semantic cues that constructs image based hybrid maps. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1316–1321. IEEE, 2010.

[12] Z. Liu and G. von Wichert. Extracting semantic indoor maps from occupancy grids. *Robotics and Autonomous Systems*, 2013.

[13] J. Mason and B. Marthi. An object-based semantic world model for long-term change detection and semantic querying. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012.

[14] A. Nüchter and J. Hertzberg. Towards semantic maps for mobile robots. *Robotics and Autonomous Systems*, 56(11):915–926, 2008.

[15] D. Pangercic, B. Pitzer, M. Tenorth, and M. Beetz. Semantic object maps for robotic housework-representation, acquisition and use. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012.

[16] J.T. Park and J.B. Song. Hybrid semantic mapping using door information. In *8th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, pages 128–130. IEEE, 2011.

[17] M. Persson, T. Duckett, C. Valgren, and A. Lilienthal. Probabilistic semantic mapping with a virtual sensor for building/nature detection. In *International Symposium on Computational Intelligence in Robotics and Automation*, pages 236–242. IEEE, 2007.

[18] A. Pronobis and P. Jensfelt. Large-scale semantic mapping and reasoning with heterogeneous modalities. In *IEEE International Conference on Robotics and Automation*, pages 3515–3522. IEEE, 2012.

[19] A. Ranganathan and F. Dellaert. Semantic modeling of places using objects. In *Robotics: Science and Systems*, 2007.

[20] R.B. Rusu, Z.C. Marton, N. Blodow, A. Holzbach, and M. Beetz. Model-based and learned semantic object labeling in 3d point cloud maps of kitchen environments. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3601–3608. IEEE, 2009.

[21] K. Sjøo. Semantic map segmentation using function-based energy maximization. In *IEEE International Conference on Robotics and Automation*, pages 4066–4073. IEEE, 2012.

[22] Tong Tao, Stephen Tully, George Kantor, and Howie Choset. Incremental construction of the saturated-gvg for multi-hypothesis topological slam. In *IEEE International Conference on Robotics and Automation*, pages 3072–3077. IEEE, 2011.

[23] Z. Tu, X. Chen, A.L. Yuille, and S.C. Zhu. Image parsing: Unifying segmentation, detection, and recognition. *International Journal of Computer Vision*, 63(2):113–140, 2005.

[24] Wei Wang, Lili Chen, Dongming Chen, Shile Li, and Kolja Kühnlenz. Fast object recognition and 6d pose estimation using viewpoint oriented color-shape histogram. In *IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2013, to appear.

[25] Felix Werner, Frederic Maire, Joaquin Sitte, Howie Choset, Stephen Tully, and George Kantor. Topological slam using neighbourhood information of places. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4937–4942. IEEE, 2009.

[26] X. Zhu, H. Zhao, Y. Liu, Y. Zhao, and H. Zha. Segmentation and classification of range image from an intelligent vehicle in urban environment. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1457–1462. IEEE, 2010.