Applying Rule-Based Context Knowledge to Build Abstract Semantic Maps of Indoor Environments

Ziyuan Liu   Georg von Wichert

Abstract— In this paper, we propose a generalizable method that systematically combines data driven MCMC sampling and inference using rule-based context knowledge for data abstraction. In particular, we demonstrate the usefulness of our method in the scenario of building abstract semantic maps for indoor environments. The product of our system is a parametric abstract model of the perceived environment that not only accurately represents the geometry of the environment but also provides valuable abstract information which benefits high-level robotic applications. Based on predefined abstract terms, such as "type" and "relation", we define task-specific context knowledge as descriptive rules in Markov Logic Networks. The corresponding inference results are used to construct a prior distribution that aims to add reasonable constraints to the solution space of semantic maps. In addition, by applying a semantically annotated sensor model, we explicitly use context information to interpret the sensor data. Experiments on real world data show promising results and thus confirm the usefulness of our system.

I. Introduction

In recent years, the performance of autonomous systems has been greatly improved. Multicore CPUs, bigger RAMs, new sensors and faster data flow have made many applications possible which seemed to be unrealistic in the past. However, the performance of such systems tends to become quite limited, as soon as they leave their carefully engineered operating environments. On the other hand, people may ask, why we humans can handle highly complex problems. Maybe the exact answer to this question still remains unclear, however, it is obvious that abstraction and knowledge together play an important role. We humans understand the world in abstract terms and have the necessary knowledge, based on which we can make inference given only partial information. As a human, if we see a desk in an office room, instead of memorizing the world coordinates of all the surface points of the desk, we will only notice that there is an object "desk" at a certain position, and even this position is probably described in abstract terms like "beside the window" or "near the door". Based on prior knowledge, we can make some reasonable assumptions, such as there could be some "books" in the "drawer" of the desk, instead of some "shoes" being inside, without opening the drawer. In our work, we aim to deploy such abilities (abstraction and inference) in the area of semantic robot mapping.

Z. Liu is with the Institute of Automatic Control Engineering, Technische Universität München, D-80290, Munich, Germany. `ziyuan.liu@tum.de`

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

II. Related Work

In general, related work on semantic robot mapping can be classified into several groups. A big body of literature focuses on semantic place labelling which divides the environment into several regions and assigns each region a semantic label, such as "office room" or "corridor". Park and Song [14] proposed a hybrid semantic mapping system for home environments, explicitly using information about doors as a key feature. Combining image segmentation and object recognition, Jebari et. al. [7] extended semantic place labelling with object detection. Based on human augmented mapping, rooms and hallways are represented as Gaussian distributions to help robot navigate in [11]. Pronobis and Jensfelt [16] integrated multi-modal sensory information and human intervention to classify places with semantic types. Other examples on semantic place labelling can be found in [3], [8] and [21].

Different from place labelling, another group of work concentrates on labelling different parts of the perceived environments with semantic tags, such as walls, floors, ceilings of indoor environments, or buildings, roads, vegetations of outdoor environments. In [12], a logic-based constraint network describing the relations between different parts is used for labelling indoor environments. Persson and Duckett [15] combined range data and omni-directional images to detect outlines of buildings and natural objects in an outdoor setting. Other examples in this category can be found in [20], [1] and [23].

Another category consists of object-based semantic mapping systems which use object as basic representation unit 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 used to represent the objects. Examples of object-based semantic mapping can be found in [19], [17], [13] and [10].

In this paper, we extend our previous work [9] using rule-based context knowledge. The work as a whole demonstrates a probabilistic method for building abstract semantic maps for indoor environments, which systematically combines data driven MCMC [22] and inference using rule-based context knowledge. Unlike semantic labelling processes, whose typical output is a map data set with semantic tags, our mapping system outputs a parametric abstract model of the perceived environment, which not only accurately represents the geometry of the environment but also provides valuable abstract information.
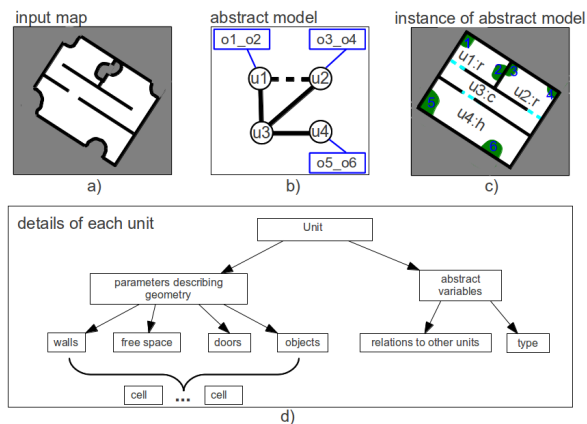
Fig. 1. Overview of our abstract semantic map model. **a)** Input occupancy grid map (gray=unknown, white=free, black=occupied). **b)** The abstract model (circle nodes=space units, blue rectangle nodes=objects, black solid edges=connected by a door, black dashed edges=adjacent without door). **c)** An instance of the abstract model (gray=unknown, white=free, black=wall, cyan=door, green=object, r=room, c=corridor, h=hall) under the assumption that each space unit has a rectangular shape. **d)** Each unit in the abstract model contains abstract variables and parameters describing its geometry (size, position and orientation).

III. AN ABSTRACT MODEL FOR SEMANTIC INDOOR MAPS

Our semantic mapping system takes a grid map (a typical result of 2D SLAM processes, e.g. [4]) of the perceived environment as input and returns a parametric abstract model of this environment which provides semantic level explanation (such as "type" and "relation") and geometrical estimation of the environment. Explanation of our model is given in Fig. 1.

Our abstract model explains indoor environments in terms of basic indoor space types, such as "room", "corridor", "hall" and so on, and we denote it as W:

$$W := \{U, T, R\}, \tag{1}$$

where $U = \{u_i | i = 1, \ldots, n\}$ represents the set of all n units. Each unit u_i has a rectangle shape, and its geometry (size, position and orientation) is represented by its four vertices. The four edges of a unit are its walls. Doors are small line segments of free cells that are located in walls and connect to another unit. Unknown cells of the input map that are located within a unit are considered as object cells. All cells within a unit that do not belong to object cells are considered as free space of the unit. $T = \{t_i | i = 1, \ldots, n\}$ is the set of type of each individual unit, with $t_i \in \{\text{room,corridor,hall,other}\}$. Here "other" indicates unit types that are not "room", "corridor" or "hall". $R = \{r_{p,q} | p = 1, \ldots, n; q = 1, \ldots, n\}$ is a $n \times n$ matrix, whose element $r_{p,q}$ describes the relation between the unit u_p and the unit u_q, with $r_{p,q} = r_{q,p} \in \{\neg\text{adjacent,adjacent}\}$. If two units share a wall, we define their relation as "adjacent", otherwise "¬adjacent". By default, we define a unit u_p is not adjacent to itself, i.e. $r_{p,p} = \neg\text{adjacent}$. In the following, we call each instance of the abstract model a "semantic world" or "world".

A main criterion for evaluating how well a semantic world W matches with the input grid map M is the posterior probability $p(W|M)$, and it is computed as:

$$p(W|M) \propto p(M|W) \cdot p(W). \tag{2}$$

Here, the term $p(M|W)$ is usually called *likelihood* and indicates how probable the input is for different worlds. The term $p(W)$ is called *prior* and describes the belief on which worlds are possible at all. In the following, we formulate task-specific context knowledge as descriptive rules in Markov Logic Networks (MLNs) [18] and show how to define the likelihood and the prior using the inference results of MLNs in a systematic way. For details on MLNs, we refer to [18].

A. Inference using rule-based context knowledge

In general, context knowledge describes our prior belief for a certain domain, such as that the ground becomes wet after it has rained. Rather than exact quantitative information, context knowledge provides advisory qualitative information for our judgements. Such information is very valuable in handling problems of high dimensionality where computation suffers due to the huge state space. In the domain of robot indoor mapping, we formulate following context knowledge:

- There are four types of space units: room, corridor, hall and other.
- Two units are either adjacent (neighbours) or not adjacent.
- The type of a unit is dependent on its geometry and size.
- In contrast to rooms, corridors have multiple doors.
- Connecting walls of two adjacent rooms have the same length.

With the help of MLNs, we formulate our context knowledge as descriptive rules in Table III. Based on these rules, query defined in Table II can be made given evidence shown in Table I. Using these rules, we try to formulate the features of certain indoor environments with rectangular space units. The choice of the rules is a problem-oriented engineering step, and the rules given in this paper serve as a good example.

predicate	explanation
$RoLi(u_p)$	Unit u_p has a room-like geometry.
$CoLi(u_p)$	Unit u_p has a corridor-like geometry.
$HaLi(u_p)$	Unit u_p has a hall-like geometry.
$MulDoor(u_p)$	Unit u_p has multiple doors.
$Adj(u_p, u_q)$	Unit u_p and u_q are adjacent.

TABLE I

DEFINITION OF EVIDENCE PREDICATES

Before we can make inference in MLNs, the evidence defined in Table I need be provided as input for MLNs, which includes geometry evidence, relation evidence and evidence on doors. To provide the first, we use a classifier that categorizes the geometry of a unit into "room-like", "corridor-like"

predicate	explanation
$Room(u_p)$	Unit u_p has the type of room.
$Corr(u_p)$	Unit u_p has the type of corridor.
$Hall(u_p)$	Unit u_p has the type of hall.
$Other(u_p)$	Unit u_p has the type of other.
$SaLe(u_p, u_q)$	Unit u_p and u_q have each a wall with the same length.

TABLE II

DEFINITION OF QUERY PREDICATES

basic features:
$Adj(u_p, u_q) \rightarrow Adj(u_q, u_p)$
$SaLe(u_p, u_q) \rightarrow SaLe(u_q, u_p)$
reasoning on type:
$HaLi(u_p) \rightarrow Hall(u_p)$
$HaLi(u_p) \rightarrow \neg Room(u_p)$
$HaLi(u_p) \rightarrow \neg Corr(u_p)$
$HaLi(u_p) \rightarrow \neg Other(u_p)$
$RoLi(u_p) \rightarrow \neg Hall(u_p)$
$CoLi(u_p) \rightarrow \neg Hall(u_p)$
$RoLi(u_p) \wedge \neg MulDoor(u_p) \rightarrow Room(u_p)$
$RoLi(u_p) \wedge \neg MulDoor(u_p) \rightarrow \neg Corr(u_p)$
$RoLi(u_p) \wedge MulDoor(u_p) \rightarrow Other(u_p)$
$CoLi(u_p) \wedge \neg MulDoor(u_p) \rightarrow Other(u_p)$
$CoLi(u_p) \wedge MulDoor(u_p) \rightarrow Corr(u_p)$
$CoLi(u_p) \wedge MulDoor(u_p) \rightarrow \neg Room(u_p)$
reasoning on $SaLe$:
$\neg Adj(u_q, u_p) \rightarrow \neg SaLe(u_p, u_q)$
$Room(u_p) \wedge Room(u_q) \wedge Adj(u_p, u_q) \rightarrow SaLe(u_p, u_q)$
$Room(u_p) \wedge Hall(u_q) \wedge Adj(u_p, u_q) \rightarrow \neg SaLe(u_p, u_q)$
$Room(u_p) \wedge Corr(u_q) \wedge Adj(u_p, u_q) \rightarrow \neg SaLe(u_p, u_q)$
$Hall(u_p) \wedge Corr(u_q) \wedge Adj(u_p, u_q) \rightarrow \neg SaLe(u_p, u_q)$
$Other(u_p) \wedge Hall(u_q) \wedge Adj(u_p, u_q) \rightarrow \neg SaLe(u_p, u_q)$
$Other(u_p) \wedge Corr(u_q) \wedge Adj(u_p, u_q) \rightarrow \neg SaLe(u_p, u_q)$
$Other(u_p) \wedge Room(u_q) \wedge Adj(u_p, u_q) \rightarrow \neg SaLe(u_p, u_q)$

TABLE III

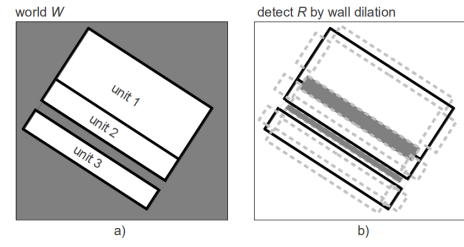CONTEXT KNOWLEDGE DEFINED AS DESCRIPTIVE RULES

Fig. 2. An example of relation detection. **a)** A semantic world W containing three units (black=wall, white=free, gray=unknown). **b)** All four walls of each unit are dilated, with dashed rectangles in light-gray representing the dilated walls. The overlap of the dilated walls is shown in dark-gray which indicates the relation of "adjacent". The overlap is detected using connected-components analysis [2]. In this example, unit 1 and unit 3 are not adjacent; unit 2 and unit 3 are adjacent; unit 1 and unit 2 are adjacent.

Given necessary evidence, we can make inference in MLNs and use the inference results to calculate the prior and likelihood. In this work, we have used the ProbCog Toolbox [6] to perform MLN inference. Currently, we use hard evidences for knowledge processing, however, our system is also able to process soft evidences, as long as the evidences are provided in the soft form.

B. Inference-based prior and likelihood design

1) Prior: According to the model definition in equation (1), the prior $p(W)$ is given by

$$
\begin{aligned}
p(W) &= p(U, T, R) \\
&= p(U|T, R) \cdot p(T, R).
\end{aligned} \tag{4}
$$

Here, $p(U|T, R)$ can be seen as a factor expressing the dependency of the geometry parameters of the underlying units (see Fig. 1-d) on the abstract terms in the MLNs. In our case, the geometry (size, position and orientation) of a unit is described by its four vertices. Furthermore, we define

$$
p(U|T, R) := \eta \prod_{p,q \in [1,2,\dots,n]} b(u_p, u_q), \tag{5}
$$

with

$$
b(u_p, u_q) = \begin{cases} e^{-\frac{d^2}{2\sigma^2}}, p(SaLe(u_p, u_q)|\text{evidence}) > \text{threshold} \\ \qquad \text{and} \quad p \neq q, \\ 1, \quad \text{otherwise,} \end{cases} \tag{6}
$$

where n is the total number of units, and d represents the length difference of the connecting walls of two adjacent units. $e^{-\frac{d^2}{2\sigma^2}}$ indicates a Gaussian function with mean at zero. $p(SaLe(u_p, u_q)|\text{evidence})$ is one of the inferences that we can make in MLNs. η is the normalization factor which ensures that $p(U|T, R)$ integrates to one. At the current stage, we assume that $p(T, R)$ follows a uniform distribution. However, it is possible to learn this distribution given proper training data.

So far, the prior $p(W)$ is defined based on the inference results of MLNs, which enforces that the semantic worlds that comply with the context knowledge have high prior probability. Note that the worlds that contradict the context

or "hall-like" according to its size and length/width ratio. The general idea of this classifier is shown in Table IV.

Relation evidence is detected based on image processing techniques: we first dilate all four walls of each unit, and then relation $r_{p,q}$ between the unit u_p and u_q is decided according to connected-components analysis [2]. An example of relation detection is depicted in Fig. 2, where R is given by

$$
R = \begin{bmatrix} \neg\text{adj} & \text{adj} & \neg\text{adj} \\ \text{adj} & \neg\text{adj} & \text{adj} \\ \neg\text{adj} & \text{adj} & \neg\text{adj} \end{bmatrix}. \tag{3}
$$

Similar to relation detection, doors are detected as small open line segments which are located on the connecting walls of two neighbour space units. Details on door detection can be found in [9].

size \ ratio	small	big
small	room-like	corridor-like
big	hall-like	hall-like

TABLE IV

THE CLASSIFIER PROVIDING GEOMETRY EVIDENCE.

knowledge are not given a zero prior probability, instead, they become less probable. The general idea of inference-based prior design is explained in Fig. 3, using a one-dimensional example.
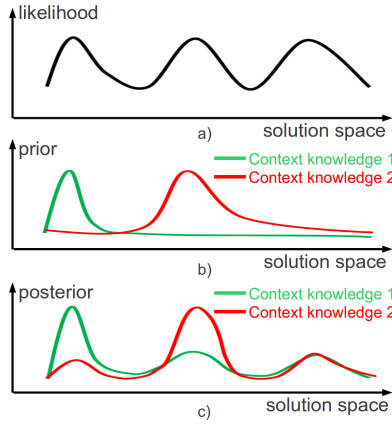

Fig. 3. The general concept of inference-based prior design illustrated using a one-dimensional example. **a)** The likelihood for different settings of models, which contains three optima. **b)** The prior distribution represented by context knowledge defined in MLNs. Different context knowledge (set of rules) represents different prior distribution (green and red). If no context knowledge is used, it is the same as implementing the context knowledge that represents a uniform distribution which does not influence the posterior, i.e. posterior is only proportional to likelihood. **c)** Corresponding posterior distributions obtained using the two prior distributions shown in figure b. By setting prior distribution by inference, we shape the posterior so that the number of optima decreases, which means, the models complying with our context knowledge tend to have high posterior probability.

2) Likelihood: Let $c(x, y)$ be the grid cell with the coordinate (x, y) in the input map M, then we define the likelihood $p(M|W)$ as follows:

$$p(M|W) = \prod_{c(x,y) \in M} \alpha(c(x,y)) \cdot p(c(x,y)|W). \quad (7)$$

Here $\alpha(c(x, y))$ penalizes overlap between units and is given by

$$\alpha(c(x,y)) = \psi^{\gamma(c(x,y))}, \quad (8)$$

with

$$\gamma(c(x,y)) = \begin{cases} \sigma(c(x,y)) - 1, & \sigma(c(x,y)) > 1 \\ 0, & \text{otherwise,} \end{cases} \quad (9)$$

where ψ is a penalization factor with $\psi \in (0, 1)$. $\sigma(c(x, y))$ indicates the number of units, to which $c(x, y)$ belongs.

In equation (7), the term $p(c(x, y)|W)$ is a semantic sensor model and evaluates the match between the world W and input map M. Essentially, $p(c(x, y)|W)$ captures the quality of the original mapping algorithm producing the grid map which is used as input in our system. For calculating $p(c(x, y)|W)$, we discretize the cell state $M(x, y)$ of the input map into three classes "occupied", "unknown" and "free", by thresholding its occupancy values. Our semantic world W contains four types of cell states, which are

- "wall": cells on the four edges of each unit.
- "object": cells that are located within a unit and are considered as non-free. These cells are detected using connected-components analysis [2].

- "free": cells that are located within a unit and do not belong to the class "object".
- "unknown": cells that are located outside all units.

In this way, our semantic sensor model $p(c(x, y)|W)$ is realized as a "3×4" look-up table.

In the real world, it is more likely that rooms and halls contain objects than corridors do, because the functionality of corridors is connecting other units, rather than placing objects. Thus, we propose to make the values of our semantic sensor model dependent on the type (decided based on the inference results in MLNs) of the underlying unit. The fundamental idea is to set the values of the semantic sensor model for the units of non-corridor types in such a way that it does not strongly penalize the mismatch between the input and the semantic world, and thus allows the existence of false positives (potential object cells). The effect of our semantic sensor model is depicted in Fig. 4.
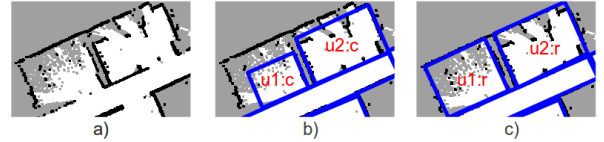

Fig. 4. Effect of our semantic sensor model. **a)** The input map. **b)** The highest likelihood solution for unit 1 and unit 2 if they are of the type "corridor". **c)** The highest likelihood solution for unit 1 and unit 2 if they are of the type "room". Note that the type of a unit is decided based on the inference results in MLNs (Table III).

IV. STOCHASTIC GENERATION OF SEMANTIC WORLDS

Given the posterior probability $p(W|M)$ defined in equation (2), we aim to obtain the *maximum a posteriori* solution W^*:

$$W^* = \underset{W \in \Omega}{\operatorname{argmax}} \, p(W|M), \quad (10)$$

where Ω indicates the solution space of semantic worlds. In order to find W^*, we use a data driven MCMC sampling technique [22]. This technique constructs a Markov chain, and each state of this Markov chain represents a semantic world. By sequentially applying transition kernels to the current world (Fig. 5) and accepting this transition by certain probability, this technique is able to efficiently draw samples from the corresponding posterior distribution. In addition to the four reversible kernel pairs that are defined in our previous work [9], we propose here a new reversible kernel "INTERCHANGE" that changes the structure of two adjacent units at the same time, without changing the total size. These structural changes are proposals that allow our system to escape local optima. Fig. 5 shows an example of the reversible MCMC kernels. More details on the realization of the used data driven MCMC process can be found in [9].

V. EXPERIMENTS AND DISCUSSION

In this paper, we extend our previous work [9] with inference using rule-based context knowledge, and the performance of our current system is shown in Fig. 6. As input, a big occupancy grid map M ("ubremen-cartesium" dataset
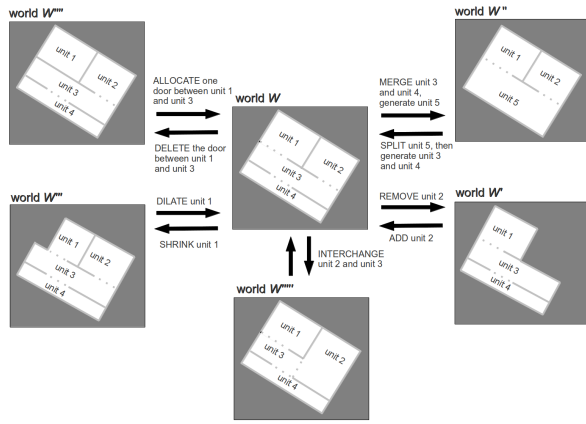
Fig. 5. Reversible MCMC kernels: ADD/REMOVE, SPLIT/MERGE, SHRINK/DILATE, ALLOCATE/DELETE and INTERCHANGE. The world W can transit to W', W'', W''', W'''' and W''''' by applying the kernel REMOVE, MERGE, SHRINK, DELETE and INTERCHANGE, respectively. By contrast, the world W', W'', W''', W'''' and W''''' can also transit back to W using corresponding reverse kernel.

[5]) of an entire floor of a building is used. This map is a big matrix ("1237×672") containing occupancy values. We classify these values as $\{occupied, unknown, free\}$ so as to generate the classified input map C_M (Fig. 6-a). Starting from a random initial guess, the semantic world W is adapted to better match the input map M by stochastically applying the kernels shown in Fig. 5. After certain burn-in time, we get the most likely semantic world W^* comprised of 17 units (each of which is represented by a rectangle) as shown in Fig. 6-c. Not only does W^* accurately represent the geometry of the input map, but also W^* is a parametric abstract model (Fig. 6-b) of the input map that provides valuable abstract information, such as adjacency, existence of objects and connectivity by doors. In addition, unexplored areas are also captured by our abstract model (marked by magenta "N"). These areas are too small to be recognized as space units but are evidence for physically existing space.

Compared with our previous work [9], our current system employs context knowledge in a systematic way, so that the input map is explained according to the underlying model structure. A performance comparison between our previous work and our current system is depicted in Fig. 7. Three high likelihood samples obtained from our previous work are shown in Fig. 7-a,b,c. They essentially represent the local maxima of the likelihood shown in Fig. 3. Although all these three results provide good match to the input map (in terms of high likelihood), they have topological defects (highlighted by magenta circles), which contradict our knowledge (low prior). In this case, all pairs of connecting walls of adjacent rooms that should have the same length are drawn in orange in Fig. 7-a,b,c. The length difference of each pair of these connecting walls results in a penalization in prior probability (equation (6)). By applying our rule based context knowledge, local maxima of the likelihood with topological defects are suppressed so that they have a low posterior probability. In this way a semantic world that has high

likelihood and high prior, i.e. high posterior (Fig. 7-d), is easily obtained by stochastic sampling. In addition, various poor local matches (highlighted by magenta rectangles in Fig. 7-a,b,c) are corrected by the semantic sensor model used in our current system.

Starting the Markov chain from the world state W^* as shown in Fig. 6-c, we show the posterior distribution obtained from our previous work (Fig. 8-a) and that obtained from our current system (Fig. 8-b) by plotting 1000 accepted samples together. Here we purposefully plot each sample (world) using very thin line. It is obvious that the underlying Markov chain obtained from our current system is more stable and converges better (smaller variance).

Fig. 8. Posterior distribution built by 1000 samples. Starting from the world state shown in figure 6-c, we plot 1000 accepted samples obtained from our previous work (figure a) and those obtained from the current system (figure b) onto the classified map . It is obvious that the underlying Markov chain converges better because of the incorporated rule-based knowledge defined in MLNs. Here, we purposefully draw each sample using very thin line.

Fig. 9 shows the performance of our current system on another data set. As can be seen, our system accurately represents the geometry of the environments captured in the input maps and provides a semantic world that explains the perceived environments with the correct topology. We evaluate our system quantitatively using the measure "cell prediction rate" (CPR), which denotes the percentage of the correctly explained cells in the manually defined region of interest (see Fig. 6-a and 9-a). The CPR of Fig. 6-c is 86.8%, and that of Fig. 9-d is 91.4%.

By modelling context knowledge in MLNs, we can assign semantic information, e.g. the type, to the data. This allows us to use a semantically informed sensor model to better explain the observations. Moreover, the used context knowledge, given by the rules, shape our prior so that unlikely configurations can be ruled out, as shown in Fig. 7.
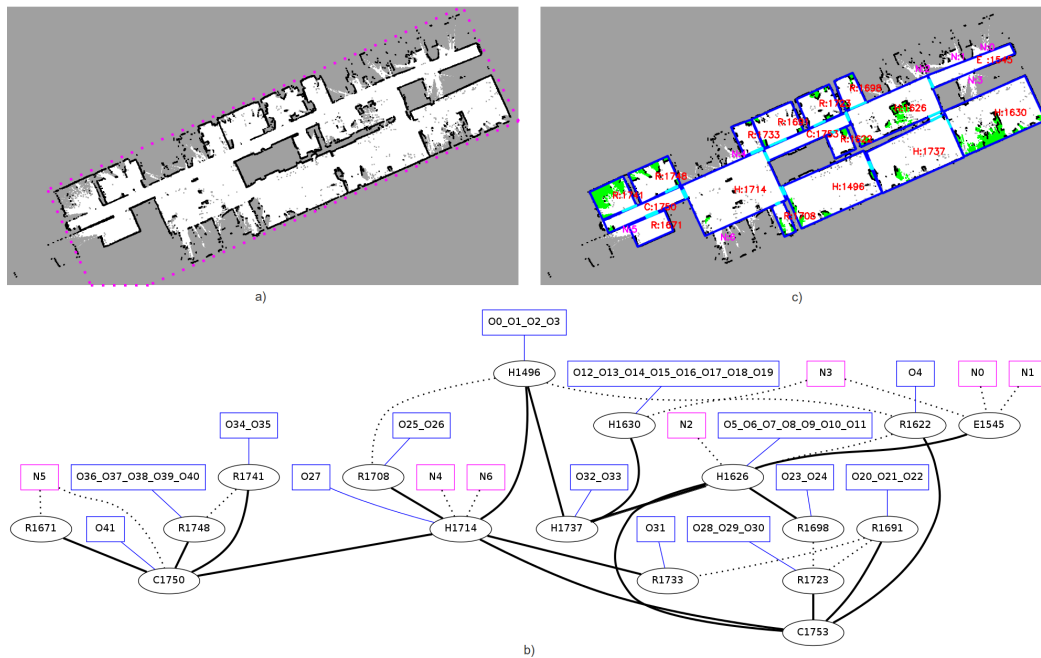
Fig. 6. The performance of our semantic mapping system. **a)** The classified input map C_M (black=occupied, gray=unknown, white=free). The manually defined region of interest (marked by magenta dashed lines) is used only for quantitative evaluation. **b)** The corresponding abstract model of W^* (ellipse node=space unit with ID, blue rectangle node=detected objects with ID, magenta rectangle node=unexplored area with ID, black solid edge=connected by door, black dashed edge=adjacent without door). **c)** The most likely semantic world W^* plotted onto the classified map C_M (black=occupied, blue=wall, gray=unknown, white=free, cyan=door, green=object). The type and ID of each unit is shown at its center (R=room, H=hall, C=corridor, E=other). Unexplored area is detected using connected-components analysis [2] and is marked by magenta "N". These areas are too small to be recognized as space unit but are evidence for physically existing space.
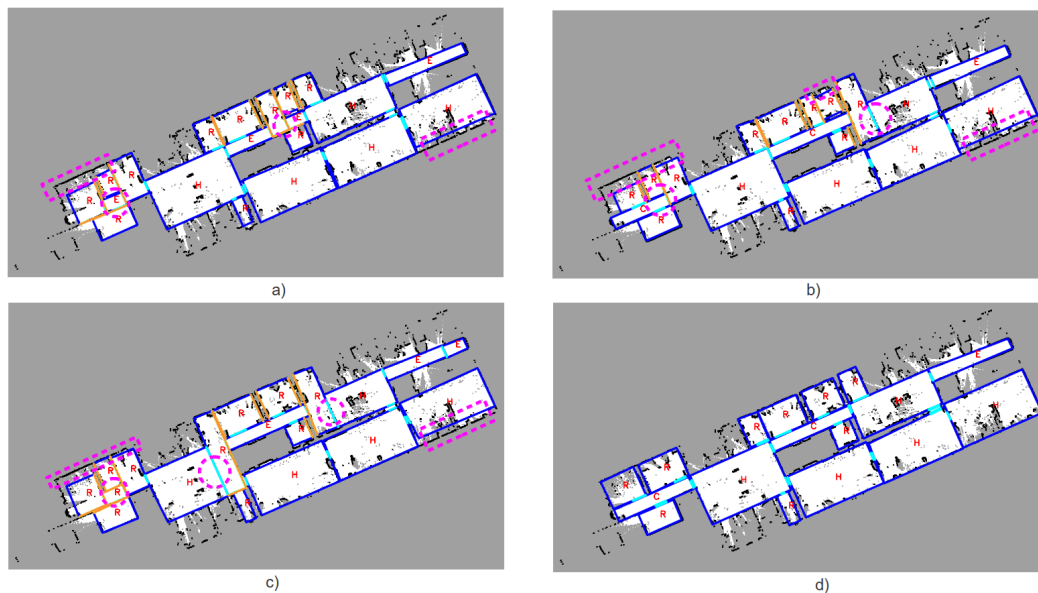

Fig. 7. Comparison of performance between our previous work [9] (figure a, b and c) and current system (figure d). Topological defects of the previous results are highlighted by magenta dashed circles, and poor local matches that are improved by semantic sensor model are highlighted by magenta dashed rectangles. Each pair of connecting walls that should have the same length are drawn in orange (figure a, b and c). Type of each unit is shown at its center.

The computational cost is strongly dependent on the size of the input map and consists of two parts, which are MCMC operations and knowledge processing in MLNs. With a single-threaded implementation on an Intel i7 CPU, the speed of MCMC operations is around 30 iterations per second for the map shown in Fig. 6. The speed of knowledge processing in MLNs depends on one hand on the tool (the software implementation of MLNs) that one uses. On the other hand, it depends on the number of optimization iterations set in the tool. In our case, we could get a satisfactory result in 5-8 seconds using the tool in [6] per processing. To analyze a grid map, we first start our system without activating

knowledge processing, only after enough context is available (e.g. coverage of the input map greater than 80%), knowledge processing is enabled to help better explain the input map. In this way, we could obtain a good result within a reasonable processing time, which is, 20 minutes for the map shown in Fig. 6.

VI. Conclusion

In this paper, we extended our previous work [9] with inference using rule-based context knowledge. Our current system demonstrates an advanced stochastic sampling process supervised by the context knowledge which is defined as descriptive rules in Markov Logic Networks. As output, our system returns a parametric abstract model of the perceived environment that not only accurately represents the environment geometry, but also provides valuable abstract information, which serves as a basis for higher level reasoning processes. By constructing the prior distribution of the semantic maps using inference results, high likelihood results with topological defects (contradiction with context knowledge) are suppressed. Furthermore, by applying a semantically annotated sensor model for likelihood calculation, we explicitly use the extracted semantic information to improve the performance of our system.
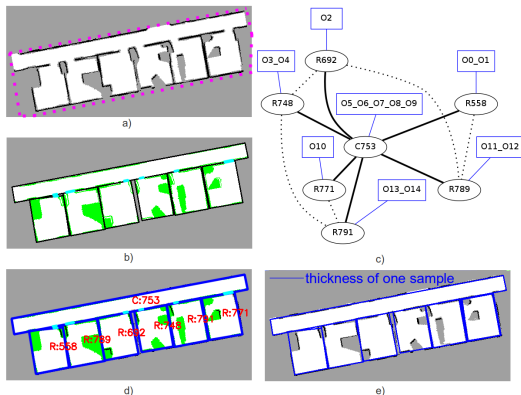

Fig. 9. The performance of our current system for another data set. **a)** The classified input map. The manually defined region of interest (marked by magenta dashed lines) is used only for quantitative evaluation. **b)** The most likely semantic world. **c)** The corresponding abstract model. Here, no unexplored area is detected. **d)** A direct comparison between the input map and the corresponding semantic world. **e)** Posterior distribution built by 1000 samples. It is obvious that the underlying Markov chain achieves stable convergence using our current system.

Acknowledgements

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] 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.

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

[4] 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.

[5] A. Howard and N. Roy. The robotics data set repository (radish), 2003.

[6] D. Jain. Probcog toolbox, http://ias.cs.tum.edu/software/probcog, 2011.

[7] I. Jebari, S. Bazeille, E. Battesti, H. Tekaya, M. Klein, A. Tapus, D. Filliat, C. Meyer, R. Benosman, E. Cizeron, et al. Multi-sensor semantic mapping and exploration of indoor environments. In *IEEE Conference on Technologies for Practical Robot Applications (TePRA)*, pages 151–156. IEEE, 2011.

[8] 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.

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

[10] 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.

[11] C. Nieto-Granda, J.G. Rogers, A.J.B. Trevor, and H.I. Christensen. Semantic map partitioning in indoor environments using regional analysis. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1451–1456. IEEE, 2010.

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] M. Richardson and P. Domingos. Markov logic networks. *Machine learning*, 62(1):107–136, 2006.

[19] 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.

[20] I. Shim, Y. Choe, and M.J. Chung. 3d mapping in urban environment using geometric featured voxel. In *International Conference on Ubiquitous Robots and Ambient Intelligence*, pages 804–805. IEEE, 2011.

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

[22] 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.

[23] D.F. Wolf and G.S. Sukhatme. Semantic mapping using mobile robots. *IEEE Transactions on Robotics*, 24(2):245–258, 2008.