IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 21, 2018, accepted September 14, 2018, date of publication October 11, 2018, date of current version November 9, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2873597

Hierarchical Semantic Mapping Using Convolutional Neural Networks for Intelligent Service Robotics

REN C. LUO, (Fellow, IEEE), AND MICHAEL CHIOU
International Center for Intelligent Robotics and Automation Research, National Taiwan University, Taipei 10617, Taiwan

Corresponding author: Ren C. Luo (renluo@ntu.edu.tw)

This work was supported in part by the Ministry of Science and Technology, Taiwan (R.O.C) and in part by the International Center of Excellence in Intelligent Robotics and Automation Research (NTU-iCeiRA) Laboratory.

ABSTRACT The introduction of service robots in the public domain has introduced a paradigm shift in how robots are interacting with people, where robots must learn to autonomously interact with the untrained public instead of being directed by trained personnel. As an example, a hospital service robot is told to deliver medicine to Patient Two in Ward Three. Without awareness of what ''Patient Two'' or ''Ward Three'' is, a service robot must systematically explore the environment to perform this task, which requires a long time. The implementation of a Semantic Map allows for robots to perceive the environment similar to people by associating semantic information with spatial information found in geometric maps. Currently, many semantic mapping works provide insufficient or incorrect semantic-metric information to allow a service robot to function dynamically in human-centric environments. This paper proposes a semantic map with a hierarchical semantic organization structure based on a hybrid metric-topological map leveraging convolutional neural networks and spatial room segmentation methods. Our results are validated using multiple simulated and real environments on our lab's custom developed mobile service robot and demonstrate an application of semantic maps by providing only vocal commands. We show that this proposed method provides better capabilities in terms of semantic map labeling and retain multiple levels of semantic information.

INDEX TERMS Intelligent robots, human–robot interaction, semantic technology, service robots.

I. INTRODUCTION

Service robots act in assistive capacities by performing repetitive or distant tasks in place of people such as domestic cleaning, educational tasks, or general guidance which aid in addressing the growing deficiency with available labor in health care. Many research topics focus on actively improving Human-Robot Interactions (HRI) by integrating robotic platforms with the ability to interpret semantic information. Integrating semantic information allows for intelligent robots to improve interactivity by enabling implicit commands or creating additional functionality [1].

Research in Simultaneous Localization and Mapping (SLAM) has arguably reached the level where robots can accurately construct a geometrically consistent global map of the environment while performing localization with respect to this map [2]. However, robots lack the ability to comprehend the environment in the same manner as a human which leads to difficulties for any human-robot interactions (HRI). Therefore, there is a need for robots to fathom the environment from human perspectives such as telling the difference between a hallway and a room. Though there already exist semantic mapping works that attempt to address this issue, many of these maps are run on either high-powered equipment, which is unsuitable for mobile systems, use a 3D CAD model-based knowledge database which are not generalizable, or use computationally expensive large convolutional neural networks (CNNs) for place categorization.

This paper approaches this mapping issue by implementing a semantic mapping framework that combines both visual and semasiological elements in a mobile service robot system running an embedded system with an integrated GPU. We utilize the GPU to run a lightweight CNN model for object recognition which has shown remarkable success in comparison to traditional methods. Using the CNN allows the robot to spend its limited resources on other tasks such as navigation. We will demonstrate the applications of semantic mapping

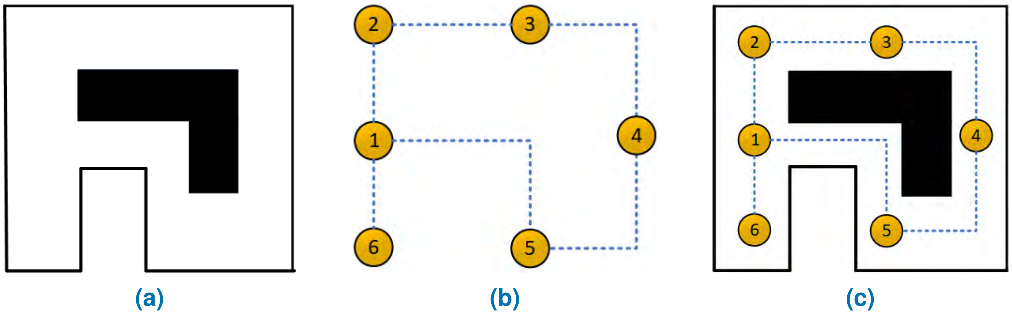


FIGURE 1. (a) shows a metric map M of size (x, y) where x and y are integers, (b) shows a topological map made of n nodes and e edges which maintain relative distances between each n_{th} node. (c) shows a hybrid metric-topological map by overlaying a topological graph over a metric map.

and intelligent behavior via semantically-based navigation for operating in human-centric environments based on audio input. In this work, we make the following contributions:

1) A custom robotics system running Robot Operating System (ROS) with semantically-cognizant capabilities using a hybrid metric-topological map.
2) Provide spatial-semantic map correlations using spatial room segmentation and a topological map.
3) Custom defined semantic hierarchical semantic structure linking abstract concepts with more tangible concepts using a lightweight CNN.
4) Enable navigation based on semantic inputs.

In the following, we will discuss related works pertaining to our semantic map in Section II. We describe our methodology and proposed approach in Section III including our robot system, and detail our experimental procedure and analysis in Section IV. We conclude this paper with Section V which describes our conclusion and future work.

II. RELATED WORK

A. METRIC-TOPOLOGICAL MAPPING PARADIGMS

As a research field, SLAM has been comprehensively developed to include map representations which can be classified into *metric*, *topological*, or *hybrid* mapping paradigms as shown in Fig. 1. We will discuss the advantages and disadvantages of each paradigm in order to lay down the foundation for our semantic map.

Metric mapping paradigms are grid-based representations that map the environment with two-dimensional or three-dimensional coordinates using visual or odometric data as shown in [3]. However, metric maps are pure geometric representations and lack the capability of efficiently storing semantic information. Topological mapping paradigms are graph-based representations that only considers places and the relations between them in the form of nodes and edges. This format allows for the storage of any key information only describes the relations between these nodes. Werner *et al.* [4] used this method for topological SLAM using particle filtering techniques and neighborhood associations.

Hybrid metric-topological mapping frameworks fuses metric and topological maps by overlaying the topological map onto the metric map. Konolige *et al.* [5] used a hybrid metric-topological map to simplify navigation by using topological maps to compute global path planning and metric maps for local path planning. This reduced the total complexity in navigation allowing for faster navigation with reactive changes to a dynamic environment. Blanco *et al.* [6] used a hybrid metric-topological SLAM system which was able to map large environments with promising accuracy. In our system, we take advantage of the utility of a metric-topological system to store semantic information generated from a CNN while retaining map accuracy.

B. CONVOLUTIONAL NEURAL NETWORKS

The usage of CNNs for visual classification tasks have been a major theme in recent academics, surpassing classical recognition methods. Pronobis and Jensfelt [7] uses Support Vector Classifiers (SVC) instead of machine learning methods and rely on a database of 3D CAD models for object recognition. Sünderhauf *et al.* [8] focused on place categorization which feeds a visual image into a large CNN to generate place labels of the environment. Goeddel and Olson [9] used a CNN architecture trained to label rooms and doors using only sensor information derived from a laser range finder. However these methods label metric maps which lead to inconsistencies or unconventional labeling dissimilar to how a person regards the environment. These CNN methods can use a variety of architectures such as VGG Net [10].

C. SEMANTIC MAPPING

Semantic mapping is a qualitative description of a robot's environment, enhancing information available to the robot to enhance navigational capabilities. Previous works have already shown some attempts at integrating semantic information. Some works such as by Galindo *et al.* [11] organized semantic information before applying to mapping methods to generate structured semantic maps. Capobianco *et al.* [12] used concept hierarchies to generate map representations and tools to generate semantic ground truths and semantic maps. Others try to procedurally generate semantic information such as using mathematical approaches to room segmentation by Bormann *et al.* [13]. There exist many different approaches to semantic mapping ranging from how low level sensory information is handled to high level feature semantic

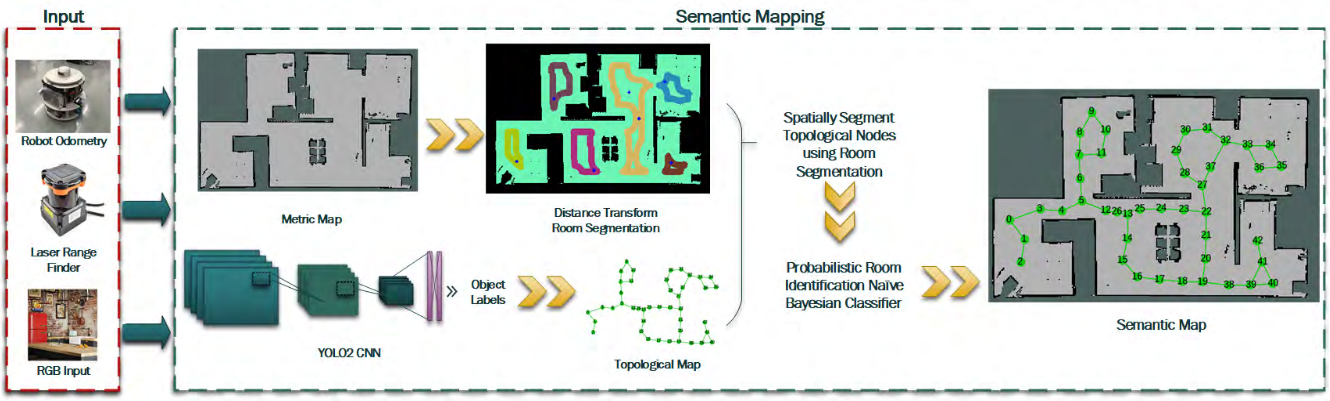

Input
Robot Odometry
Laser Range Finder
RGB Input
Semantic Mapping
Metric Map
Distance Transform Room Segmentation
YOLO2 CNN
Object Labels
Topological Map
Spatially Segment Topological Nodes using Room Segmentation
Probabilistic Room Identification Naïve Bayesian Classifier
Semantic Map

FIGURE 2. Overall system process flowchart for semantic mapping.

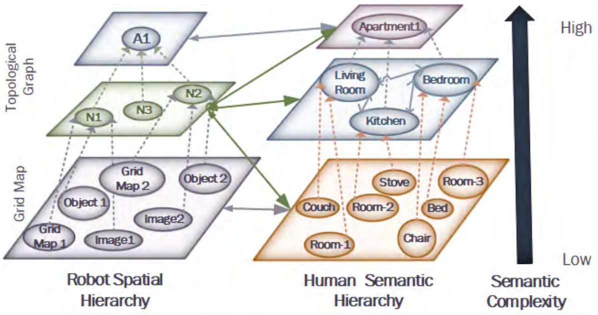
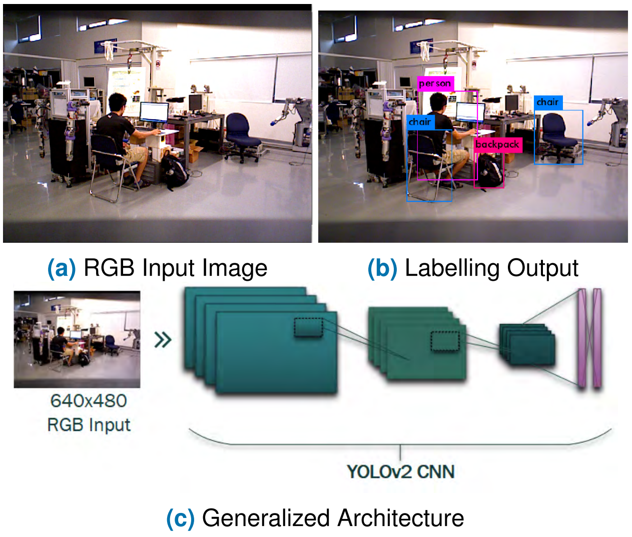


FIGURE 3. Our hierarchical semantic structure is based on semantic abstraction levels. The tiers are described in order from ascending order, spatial occupancy map, topological map, supernode for the Spatial Hierarchy. The semantic hierarchy organizes items based on semantic complexity (lower is less complex).



(a) RGB Input Image (b) Labelling Output

(c) Generalized Architecture

FIGURE 4. (a) shows a detection image as input into the YOLOv2 CNN, (b) shows the CNN output with attached labels (c) shows the general YOLOv2 CNN architecture.

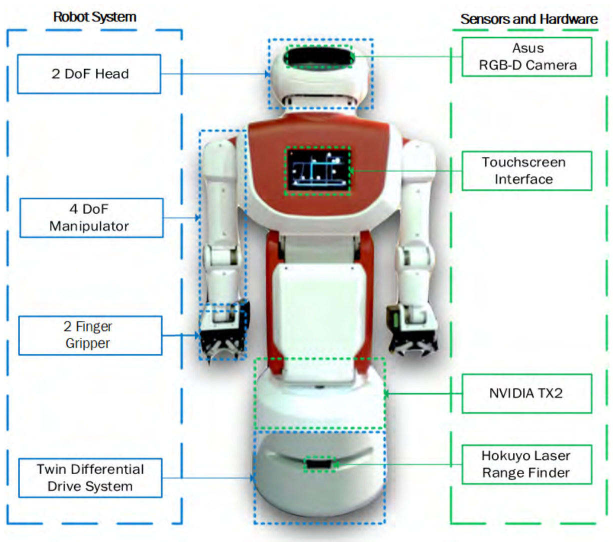


FIGURE 5. A system diagram of our service robot (Renbo-S) with a breakdown of available equipment and sensors.

feature organization. An example would be how an environment is perceived, from singular image scenes to large scale map representation. In our work, we use metric-topological maps and we generate our own semantic organizational structure based on the adage of 'form follows function' similar to the approach in [14], borrowing the concept that rooms can be identified from objects found in it. The proposed method is described in detail in Section III.

III. PROPOSED APPROACH

Our semantic mapping framework is shown in Fig. 2 using the items described in the following subsections. We store semantic information generated from a CNN and store it in the corresponding topological node in the hybrid map. We spatially segment out areas and cluster topological nodes together based on this spatial segmentation. We extract out the semantic information from the topological node cluster to determine room labels (higher semantically complex information) through a probabilistic classification.

A. HYBRID MAP STRUCTURE

For a rigorous definition, we denote the entire physical environment world as $W = < Q, P, \tau >$ where Q is the set of entities or items, P is the set of poses in 6 dimensions, and τ is the set of sematological attributes or traits. Each entity $q \in Q$

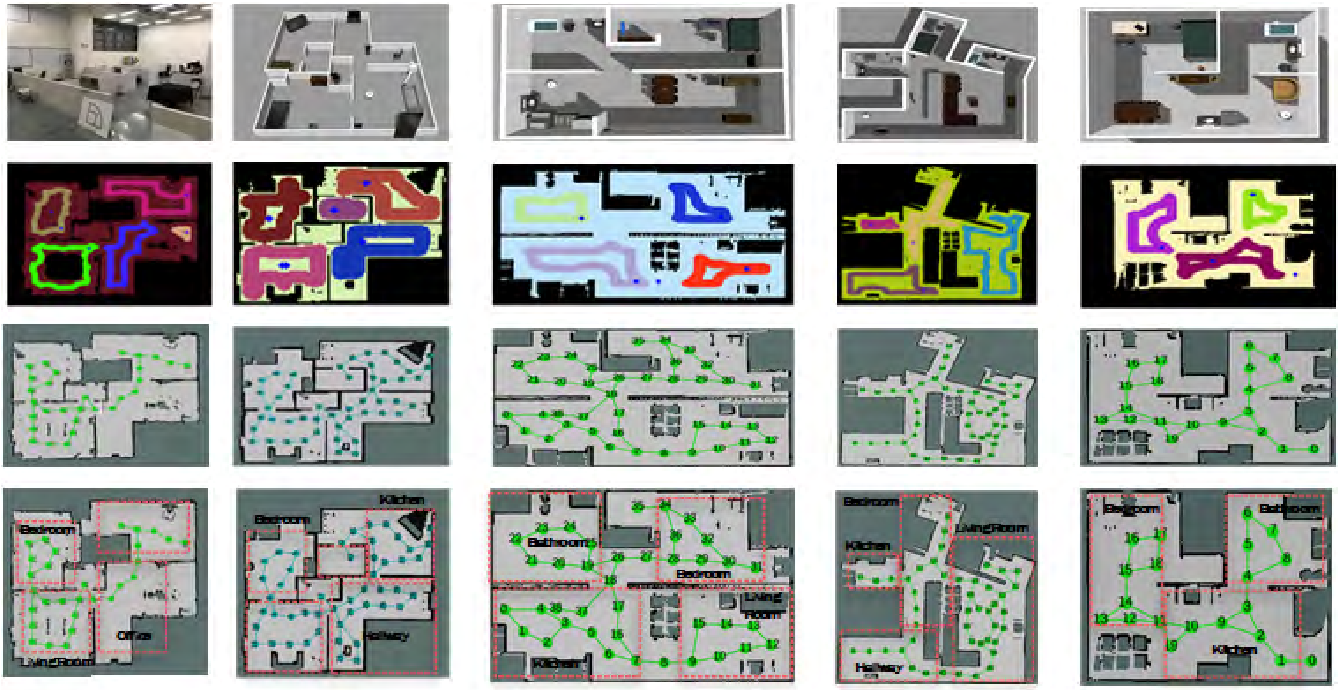


FIGURE 6. The first row describes the environment. The second row shows the final room segmentation result as the service robot finishes exploring. The third shows the output hybrid metric-topological map. The fourth row shows the final semantic mapping results after aggregating spatial semantic information from rows two and three. We manually label the output results and show the results.

is associated with a single pose $p \in P$ and a subset of characteristics $c : q \to 2^{\tau}$. This takes form as a 2D occupancy grid map M with an overlaid topological graph $T(N, E)$ where T is an undirected topological graph consisting of the Node set N and Edge set E. For each node $n_i \in N$ contains navigational waypoints or semantic trait t_k where i is the index of the node and k is the trait type respectively. Both topological graph T and the occupancy map M are simultaneously generated and store semantic information and trait t_k as generated.

B. HIERARCHICAL SEMANTIC ORGANIZATION

For any world W, it is imperative to organize information defined as a concepts that acts as a super-categorical for all subordinate concepts and connects any related concepts as a group, field, or category together. We organize every required semantic concept $t \in \tau$ that our robot must be aware in terms of size and abstraction level in a hierarchical format as shown in 3. Tangible objects which can be directly labeled by sight such as doors and chairs are considered as the lowest tiers in both the spatial and semantic hierarchies. Abstract concepts such as rooms and room labels such as kitchens, offices, or bedrooms are classified as the next abstraction tier. Theoretically, this hierarchical structure can be used to describe all semantic objects with relation to each other from a first principles approach with each subsequent tier describe more abstract concepts.

C. MACHINE LEARNING FOR OBJECT RECOGNITION

We utilize CNNs which has been used in other semantic mapping works via place categorization to generate labels. However, we approach semantic information generation differently. Instead of identifying the entire scene via place categorization CNN, we use the CNN only for object recognition and store this information in a topological node. This allows for a smaller and compact CNN architecture to be implemented. This removes the need to specifically identify scenes based on a single visual input, but instead provide multiple inputs based on determined labels which provides more information about a given area then what can be done by place categorization CNN. We implement the YOLOv2 [15] architecture as our CNN model due to its accuracy and comparatively small size and show a sample image in Fig. 4 which is capable of running at 10FPS. We use household objects from the Microsoft COCO Dataset [16] to train our object recognition CNN.

D. SPATIAL-TOPOLOGICAL ROOM SEGMENTATION

To correlate relevant basic spatial information such as room size with the relevant semantic information in the correct topological node n, it is necessary to group the nodes $n \in N$ in the topological graph T in a room M_r where r is a identified room. We perform a spatial room segmentation to identify regions of interest in the metric map M which belong to a room so that topological nodes can be spatially segmented out. This not only allows for segmentation of the correct topological nodes n_i but simultaneously provides additional geometric information such as the spatial size of the room which can aid in classification. This process is performed as shown in the following. We preprocess our map by using (1), and threshold the grid map which consists of $M(x, y)$ where x and y are coordinates in map M and δ is the threshold value.

$$M_t(x, y) = \begin{cases} 1 \ if \ M(x.y) \leq \delta \\ 0 \ if \ M(x, y) \geq \delta \end{cases} \tag{1}$$

We take M_t and apply an erosion morphological transformation to eliminate small noise and mapping errors as shown in (4). (3) and (4) show the basic formula for Erosion and Dilation operators. This relies on an image A with structuring element B that exist in Euclidean space E.

$$A \ominus B = z \in E | B_z \subseteq A \tag{2}$$

$$A \oplus B = \bigcup_{b \in B} A_b \tag{3}$$

$$A \circ B = (A \oplus B) \ominus B \tag{4}$$

With a M_{clean} map devoid of noise, we apply the distance transformation shown in (5) using the Euclidean distance function shown in (6)

$$d(p, q) = d(q, p) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \tag{5}$$

where p and q are two points in the grid map M with coordinates (x, y).

$$M_{dist} = d_f(p) = min_{q \in M}(d(p, q) + f(q)) \tag{6}$$

With a distance transformed map M_{dist}, we clean up the map by normalizing and thresholding the map which is respectively shown in (7) and (8).

$$M' = \frac{X - X_{min}}{X_{max} - X_{min}} * T_{max} \tag{7}$$

$$f(r_i) = \begin{cases} 1 \ if \ M(x.y) \leq \delta \\ 0 \ if \ M(x, y) \geq \delta \end{cases} \tag{8}$$

We use an energy function to optimize for the rooms areas and cross-reference the room areas with the topological node $n \in T$ using a Breadth-First Search graph traversal shown in Algorithm 1. Cross-referencing allows for T_n subgraphs to be made which are then used to identify room labels using the aggregated semantic information collected in each node.

Algorithm 1 Breadth-First Search

```
Input ← Graph $T$ and node $n \in T$
Output ← All reachable nodes from $n$ as discovered

**Let** Q be queue
Q.enqueue($n$)   ▷ insert $n$ in Q until all neighbors marked

mark $n$ as visited

**while** Q is not empty **do**
    v = Q.dequeue( )
    **for** all neighbours $w$ of $v$ in $T$ **do**
        **if** $w$ is not visited **then**
            Q.enqueue(w)          ▷ Store $w$ in Q
            mark $w$ as visited
```

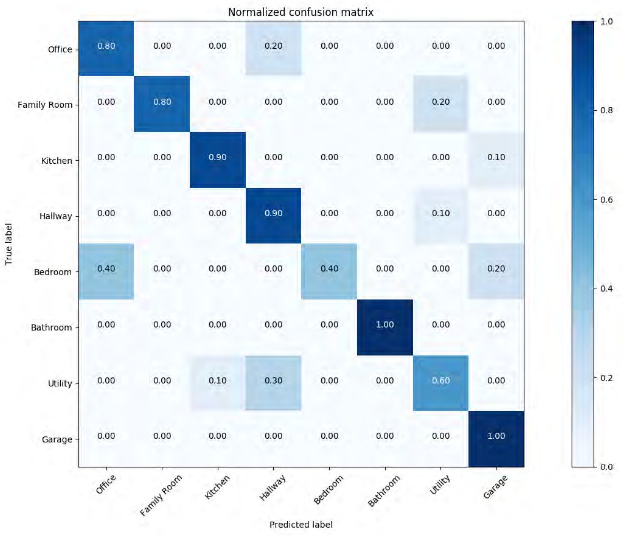


FIGURE 7. Confusion Matrix for Semantic Classification given semantic information stored in topological nodes with approximately 80% accuracy.

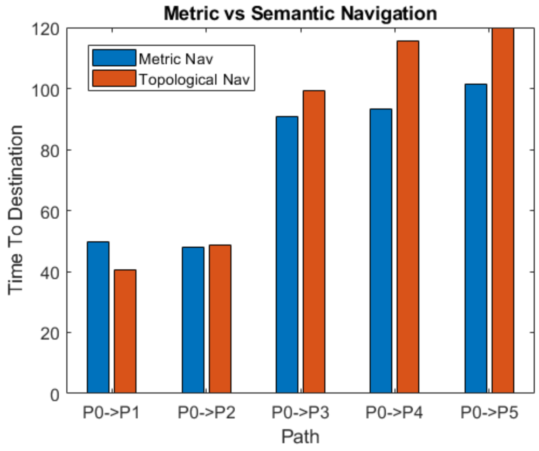


FIGURE 8. Time trials of Traditional vs Semantic Navigation. Each trial is performed 5 times with points defined in Fig. 8.

E. SEMANTIC CLASSIFICATION

Both spatial and topological information is collected and binarized into categories for classification. Data is binarized to remove instances of multiple object recognition instances of the same object.

Since environments are dynamic and constantly changing, we assume each object label C_k is conditionally independent of each room class x. The quantity of objects found can also be a factor in determining the correct room label. Using these assumptions, we implement a Multivariate Bernoulli Naive Bayesian model shown below.

$$p(x|C_K) = \prod_{i=1}^{n} p_{k_i}^{x_i} (1 - p_{ki})^{(1-x_i)} \tag{9}$$

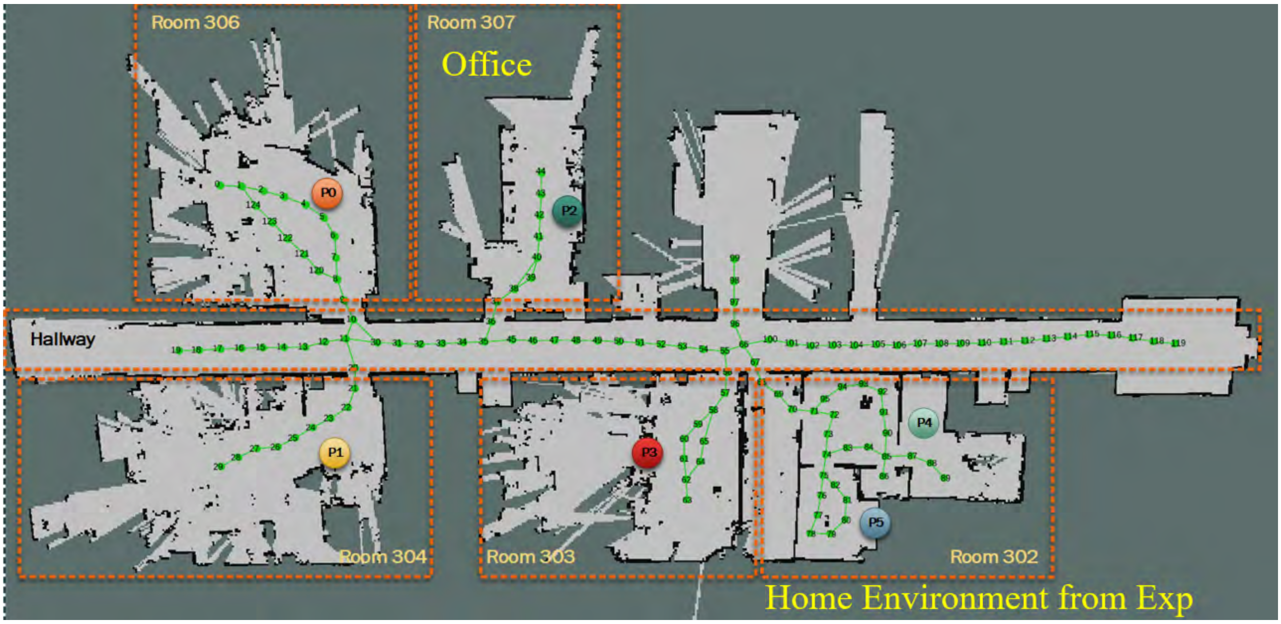


FIGURE 9. The output semantic map with areas labelled by hand. Due to a lack of household objects, only Room 307 (Office) and Room 302 (Lab Room Environment from previous experiments) are properly classified.

TABLE 1. List of semantic room classes.

Semantic Room Classes	
Office	Kitchen
Living Room	Garage
Hallway	Utility Closet
Bedroom	Bathroom

where p_{ki} is the probability of the given binary object classes C_k to determine class x. The correct room classification can be determined by feeding the detected object labels from each subgraph G_s into this Bayesian model. We categorize object labels trained in the CNN model and room characteristics into 8 room classifications normally found in a residential house as shown in Table 1.

F. ROBOT SYSTEM

Our robotic system ('Renbo-S') is shown in Fig. 5 which was developed in-house by our own NTU lab. It uses a differential drive system with a mobile platform actuated with two active wheels and 4 passive caster wheels with a maximum speed of 0.5m/s. It is equipped with two 4 Degree of Freedom (DoF) manipulator arms for basic service tasks with a 2 DoF head (pitch, yaw) for expanded visual range using an ASUS Xtion sensor. Our robot runs on a NVIDIA Jetson TX2 platform which is equipped with a NVIDIA Pascal GPU and uses ROS Kinetic under a Ubuntu 16.04 LTS system.

IV. EXPERIMENTAL RESULTS

We divide our experiments into two sections. In the first, we analyze our semantic map framework in terms of semantic labeling accuracy and mapping accuracy. In the second section, we show how semantic navigation can be performed by applying our semantic map.

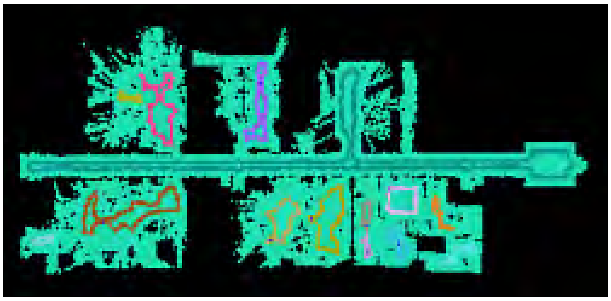


FIGURE 10. Real Room Segmentation Results.

A. SEMANTIC MAPPING RESULTS

For testing our semantic mapping method, we use several simulated home environments ranging from irregular layouts to 1 bedroom studio apartments. In addition, we use a real and simulated version of our lab room to further validate our process. We navigate our service robot and allow it to explore the entire environment. Due to several restrictions such as limited CNN training data on certain objects or real restrictions (such as finding a toilet to move into our real environment), we remove the room classification labels for Garage and Utility Closet. In each simulated environment, we include a kitchen, living room, and bedroom. Dependent on the actual floor layout, we include other labels such as Hallways or Office where applicable. The complete semantic mapping results are shown in Fig. 6.

The second row in Fig. 6 shows how our room segmentation performs with and shows results that is able to reasonably segment out rooms. There are occasional artifacts in the map

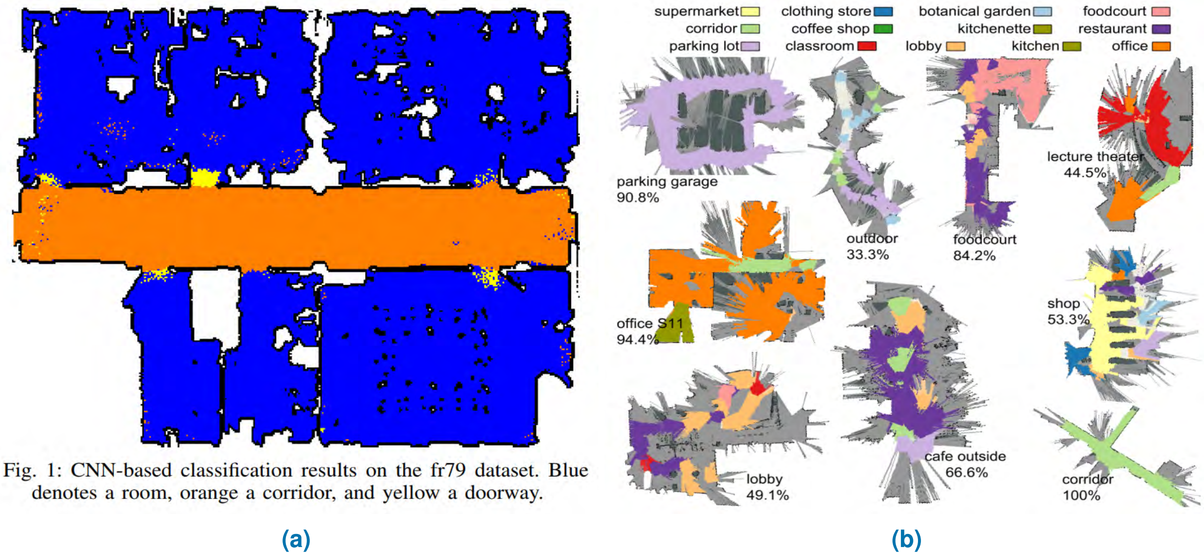


FIGURE 11. Semantic Work by Sünderhauf *et al.* to label areas based on visual input with a ConvNet [17]. (a) Sample Semantic Map from [9]. (b) Set of Semantic Map Works provided by [17].

which cannot be removed such as large furniture which can cause some variation in accuracy mapping accuracy. It can be observed that our method is capable of segmenting out rooms with relative accuracy though smaller rooms may be incorrectly clustered with larger adjacent rooms based on its size. We show in Fig. 7 that our semantic classification provides decent results by testing with 80%. This accuracy rating is only for determining if a correct room label has been generated properly despite some rooms having been incorrectly clustered into others. To analyze the entirety of our labels including the excluded garage and utility closet labels, we collect over 80 samples of clustered information and show the results in Fig. 7. It is observed that there are some trends in our classification model which may require additional data since the 'Bedroom' label and the 'Office' label show lower classification scores compared to others. This could be due to the wide variety of items that are present in both rooms.

B. SEMANTIC NAVIGATION RESULTS

We reapply our semantic mapping method on the entire floor of our lab and show the result in Fig. 8. Since our CNN is trained for residential environments, we limit the semantic room classifications to the mock-up home environment as shown in the first column of Fig. 6. We use Google Speech API to provide vocal input and navigation with audio input provides similar navigation times compared to traditional methods in Fig. 8. Traditional navigation is performed by giving the final geometric coordinates stored in the final topological navigational goal.

C. DISCUSSION AND ANALYSIS

If we use utility and practicality as criteria for comparing semantic maps, the proposed semantic map is more practical based on the following reasons in comparison to related works. In both works provided by Süderhauf *et al.* or Goëddel et al, there are semantic labeling inaccuracies which are found in the metric map. In Fig. 11a, while there are not many, small particles representing the wrong labels are present in the map as shown circled. While this may visually prove to be a small issue, in the case of semantic navigation, it provides very difficult to navigate according to computer since there exist numerous possibilities. In Fig.11b, this work chooses to semantic label a map based on visual input with a CNN. This CNN creates a feature map which directly produces a place classification label and results in comparatively good accuracy but the labels are mismatched and overlapped in some areas which contrasts with conventional human standards.

Our method prevents this mislabeling from occurring by focusing on classifying the occupancy map according to room segmentation methods. These predicted semantic labels are stored in nodes which can be easily accessible for robots to perform autonomous functions using the proposed work. We break up this method by using a CNN to create features maps to detect objects and labels and later cluster these labels based on spatial location. This allows us to achieve similar results and eliminate mismatch errors and attain semantic place labeling similar to human standards with an 80% semantic mapping accuracy. Despite this, we are able to retain extra semantic information at multiple semantic conceptual levels while using a structurally smaller CNN allowing for the implementation on an actual service robot. An example of multi-semantic level awareness would be if the given verbal command was given to a service robot running the proposed map. "Go to the Living Room and get me the TV Remote". The robot would be capable of understanding the semantic labels termed 'Living Room' and a object with a given label 'TV Remote' which is shown in the Semantic Navigation. In addition, by using a topological map, we reduce the total required computations to determine the target goal.

Despite the improvements mentioned, there are some limitations in our work. Semantics labels definitions are colloquial in nature and may change over time or regions which can affect semantic classification. This necessitates a more comprehensive knowledge database required for better semantic intelligence which may take the form of an online database. Most semantic errors were generated from incorrect spatial room segmentations which can be affected by the total space occupied by objects and irregularly shaped rooms.

V. CONCLUSION

In this paper, we developed a semantic mapping framework utilizing spatial room segmentation, object recognition trained CNNs, and a hybrid map in our lab's custom developed service robot running on a NVIDIA TX2 platform using a first principles approach. This is accomplished by extracting, aggregating, and binarizing geometric and topological information via spatial room segmentation methods to classify semantic room labels using a Bayesian Classifier with 80% accuracy. Other method differs from others in that we aggregate information using room segmentation to better classify areas whereas other works use large CNNs to immediately identify areas based on a singular images. Our method improves classification accuracy while simultaneously providing more intuitive mapping and additionally demonstrate that semantic navigation is possible using only audio input. Future work will include improved room segmentation methods and a more comprehensive semantic organization method.

REFERENCES

[1] R. C. Luo and C.-J. Chen, "Recursive neural network based semantic navigation of an autonomous mobile robot through understanding human verbal instructions," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 1519–1524.

[2] C. Cadena *et al.*, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1309–1332, Dec. 2017.

[3] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Robot.*, vol. 3, no. 5, pp. 1255–1262, Oct. 2017.

[4] F. Werner, F. Maire, H. Choset, J. Sitte, S. Tully, and G. Kantor, "Topological SLAM using neighbourhood information of places," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2009, pp. 4937–4942.

[5] K. Konolige, E. Marder-Eppstein, and B. Marthi, "Navigation in hybrid metric-topological maps," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Automat. (ICRA)*, May 2011, pp. 3041–3047.

[6] J.-L. Blanco, J.-A. Fernandez-Madrigal, and J. Gonzalez, "A new approach for large-scale localization and mapping: Hybrid metric-topological SLAM," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Automat. (ICRA)*, Apr. 2007, pp. 2061–2067.

[7] A. Pronobis and P. Jensfelt, "Large-scale semantic mapping and reasoning with heterogeneous modalities," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Automat. (ICRA)*, May 2012, pp. 3515–3522.

[8] N. Sünderhauf *et al.*, "Place categorization and semantic mapping on a mobile robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Automat. (ICRA)*, May 2016, pp. 5729–5736.

[9] R. Goeddel and E. Olson, "Learning semantic place labels from occupancy grids using CNNs," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 3999–4004.

[10] K. Simonyan and A. Zisserman. (Sep. 2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[11] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J. A. Fernandez-Madrigal, and J. Gonzalez, "Multi-hierarchical semantic maps for mobile robotics," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Aug. 2005, pp. 2278–2283.

[12] R. Capobianco, J. Serafin, J. Dichtl, G. Grisetti, L. Iocchi, and D. Nardi, "A proposal for semantic map representation and evaluation," in *Proc. Eur. Conf. Mobile Robot. (ECMR)*, 2015, pp. 1–6.

[13] R. Bormann, F. Jordan, J. Hampp, W. Li, and M. Hägele, "Room segmentation: Survey, implementation, and analysis," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, May 2016, pp. 1019–1026.

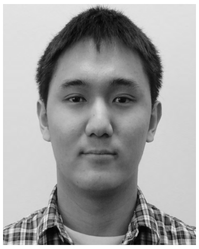
[14] H. Zender, O. M. Mozos, P. Jensfelt, G.-J. M. Kruijff, and W. Burgard, "Conceptual spatial representations for indoor mobile robots," *Robot. Auto. Syst.*, vol. 56, no. 6, pp. 493–502, 2008.

[15] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.

[16] T.-Y. Lin *et al.* (May 2014). "Microsoft COCO: Common objects in context." [Online]. Available: https://arxiv.org/abs/1405.0312v3

[17] N. Sünderhauf, S. Shirazi, F. Dayoub, B. Upcroft, and M. Milford, "On the performance of convnet features for place recognition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep./Oct. 2015, pp. 4297–4304.

REN C. LUO (M'83–SM'88–F'92) received the Dipl.-Ing and Dr.-Ing degrees in electrical engineering from the Technische Universitaet Berlin, Germany. He is currently serving as the Chief Technology Officer of the FFG-Fair Friend Group. He is also a Chair Professor at National Taiwan University. He also served two-terms as the President of National Chung Cheng University. He also served as a Tenured Full Professor of NCSU, Raleigh, USA, and a Toshiba Chair Professor at the University of Tokyo, Japan. His research interests include intelligent robotic systems, multisensor fusion and integration, and 3-D printing manufacturing. He has authored more than 500 papers on international refereed journals and refereed international conferences and international patents on these topics. He also served as the Adviser of the Ministry of Economic Affairs and Science and the Technical Adviser of Prime Minister's Office in Taiwan. He is a fellow of the IET. He served as the President of the IEEE Industrial Electronics Society. He is currently serving as the Editor-in-Chief of the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS.



MICHAEL CHIOU received the B.A.Sc. degree in mechanical engineering (robotics and bioengineering) from the University of Toronto. He is with the College of Electrical Engineering, National Taiwan University, under Professor R. C. Luo at the International Center of Excellence in Intelligent Robotics and Automation Research. His research interests include the application of robotics, computer vision, and machine learning for semantic SLAM applications.