

# A Novel Semantic SLAM Framework for Humanlike High-Level Interaction and Planning in Global Environment

Sumaira Manzoor, Sung-Hyeon Joo, Yuri Goncalves Rocha, Hyun-Uk Lee, Tae-Yong Kuc  
College of Information and Communication Engineering,  
Sungkyunkwan University, South Korea  
{sumaira11, sh.joo, yurirocha, zlshvl36, tykuc}@skku.edu

## Abstract

In this paper, we propose a novel semantic SLAM framework based on human cognitive skills and capabilities that endow the robot with high level interaction and planning in real-world dynamic environment. Two-fold strengths of our framework aims at contributing: 1) A semantic map resulting from the integration of SLAM with the Triplet Ontological Semantic Model (TOSM); 2) Human-like robotic perception system that is optimal and biologically plausible for place and object recognition in dynamic environment proposing semantic descriptor and CNN. We demonstrate the effectiveness of our proposed framework using mobile robot with Zed camera (3D sensor) and a laser range finder (2D sensor) in real-world indoor environment. Experimental results demonstrate the practical merit of our proposed framework.

## 1 Introduction

Building the autonomous mobile robot with human-like intelligence for semantic map construction and cognitive vision-based perception are two the most significant challenges for long-term planning and high-level interaction in indoor environment.

The problem to determine the appropriate method for building and maintaining the map that encodes both casual and world knowledges has become an active research area in the robotics. Many studies in the last decades have focused on spatial representation of the environment for building metric, topological and appearance-based maps. However, semantic mapping of environment for the robots has not been as intensively studied. The information provided by the conventional mapping approaches assists only in robot navigation while qualitative information about the structure of environment for task planning is not generated. For instance, metric map that contains geometric representation of the environment provides shape of the room without any semantic understanding to indicate whether it is office or lecture room. Our proposed framework tackles this issue by constructing the map that combines spatial representation with semantic knowledge of environment and provide autonomous navigation to robot for perform high-level task without human intervention in global dynamic environment.

The semantic interpretation of the environment also plays an essential role to improve the perception ability of the robot for performing real-world operations such as object and place recognition in more reliable and intelligent

manner. Nowadays, approaches for robotic perception range from traditional computer vision using handcrafted features to advanced deep learning with convolutional neural network or combination of both. However, these artificial vision algorithms have practical limitations to process in real time [boh17]. Therefore, biologically plausible algorithms combined with analogies of artificial perception are getting the attention. Our proposed framework handles the current challenges by developing the effective solution that enables the robot with the potential of human-like vision for recognizing the objects and places using semantic perception.

The primary goal of our novel semantic framework is twofold for developing semantic perception system and endowing the robot to incrementally build a consistent semantic map while simultaneously determining its location within map.

Our proposed semantic SLAM framework makes an original contribution to three important research areas in robotics with the following characteristics:

- Human-like brain GPS system for building semantic maps with emphasis on qualitative description of robot's surrounding
- Human cognition based TOSM with deeper domain knowledge acquired by semantic, topological and geometric properties of the objects for providing the robot higher degree of autonomy and intelligence.
- Bio-inspired semantic perception system combined with object and place recognition that allows the robot to relate what it perceives using semantic descriptor

This paper is organized as follows. In Section II, we provide an extensive literature review of semantic mapping, semantic SLAM and perception system for autonomous mobile robot. In Section III, we explain the key features of our proposed framework with complete details of major components of TOSM and recognition model. In Section IV, we examine the significant effects of our proposed framework in simulated environment as an illustration of its contents. Finally, we conclude our work with future direction in Section V.

## 2 Related Work

We focus our review on studies of three major concepts, which we consider are the most closely related to our work: a) semantic SLAM b) ontology c) semantic perception for object and place recognition d) semantic descriptor.

### 2.1 Semantic SLAM

This section, gives the understanding of SLAM, explain semantic SLAM structure, its concepts and related work in this area.

#### A. Semantic Mapping

In the last few years, embedding the map with semantic information has become an active research area with the motivation of human-like robot interaction and understanding of the environment. High-level features in semantic map are used to model the human concepts about the objects, places and relationship between them [Capobianco15]. Semantic mapping has recently become the center of attraction in research community which divides the semantic mapping approaches into three groups based on object, appearance and activity [Pendleton17]. Object based semantic mapping [Vasudevan08] methods depends on the occurrence of key objects to perform object recognition and classification tasks by semantic understanding of environment. Appearance based semantic mapping approaches take sensor readings and interpret them for constructing semantic information of the environment. Some studies use geometric features [Burgard07] and vision fused with LIDAR data for world understanding and classification [Nüchter08] task. The activity based semantic mapping [Xie13] techniques use information of external activities (e.g. sidewalk verses roads) around the robot for semantic understanding and contextual classification of environment. These techniques are at their formative stage compared to other two semantic mapping methods.

#### B. Semantic SLAM: Concepts

The large number of concepts and relationship among them in real-world environment lead to several task-driven decisions which depends on the level of semantic organization and context of environment in which robot

performs its task. Literature review shows two major concepts of constructing semantic relationships [cadena16] based on the details and organization. The detail of semantic concept significantly affects the complexity of the problem at different levels. For example, a robot needs only coarse categories such as rooms, doors and corridors to perform a task “going from 1st room to 2nd room” while for the other task “pick up the glass” it needs to know finer categories such as table, glass or any other object. The semantic concepts are not limited because a single entity or object in real-world environment has many properties or concepts. For example, “moveable” and “sittable” are the properties of a chair while “movable” and “unsittable” are the properties of a table. Both table and chair have same class “Furniture”. However, they share “moveable” property with different usability. So, this multiplicity of concepts is handled by Flat or hierarchical organization of properties

### Semantic SLAM: Object/ Place Recognition

Semantic to the SLAM is included by using human-spatial concepts into the maps. Humans locate themselves by object centric concepts instead of metric information and they use reference points rather than global coordinates. The initial research into semantic mapping uses direct approach [Lowry16] with metric map segmentation built by traditional SLAM system into semantic concepts. An early work in [Sabourin10] develops a system for scene understanding via semantic analysis using image segmentation techniques and the SLAM algorithm is driven by object recognition using human spatial concepts. The work shows that semantic concepts are organized in in coarse to finer manner for indoor environment. An online semantic mapping framework [Pronobis12] of indoor environment combined with object observations such as shape, size, room’s appearance that is built using three layers of reasoning to address the problem of detection and learning of novel properties and room categories for fully self-extendable semantic mapping. Data association problem also exists in metric and semantic SLAM when building a map of environment with large number of objects of the same or different class and scales. This problem is addressed in [Bowman17] by coupling geometric and semantic observations and taking the advantage of object recognition for providing meaningful scene interpretation with semantically labeled landmarks.

## 2.2 Ontology

In recent years, reducing the semantic gap using ontologies has been studied by many researchers. An early study [Durand07], has introduced an object recognition approach based on ontology and assigned the semantic meaning to objects by matching process between concepts and objects. The work in [Ji12], handles the robot task planning issues in domestic environment at the high symbolic level by combining classical AI approaches with semantic knowledge representation. Its framework is based on semantic knowledge ontology to represent robot primitive actions and description of environment. A study in [Riazuelo15], described the RoboEarth project using knowledge-based system to provide web and cloud services to multiple robots. Its semantic mapping system is based on visual SLAM mapping and ontology to describe the concepts, relations in maps and objects. A robotic system with advanced abilities leads to the complexity in its software development. A case study presented in [Saigol15] addresses this issue using an ontology as the central data store to process all information and showed that knowledge-base makes the robotic system easier to develop, modify and understand. In the last few years, a variety of approaches have been investigated to process the sensory information in dynamic world. Among them, OnPercept [Azevedo18] is a recent approach that is based on cognitive ontology for performing the HRI tasks by modeling the sensory information. A study [Lee18], proposes context query-processing framework using spatio-temporal context ontology for enabling the indoor service robots to adapt the dynamic change from the sensors in highly complex environment.

## 2.3 Perception

Perception system endows the robot to perceive and reason about its environment. The autonomous mobile robot can perform its complex tasks such as object and place recognition, collision avoidance, task planning, decision making, mapping, dynamic interaction, localization, and intelligent reasoning with high accuracy if perception information is carefully processed. A recent study has [Sünderhauf18] highlighted the fact that robotic perception is different from conventional computer vision perception because in computer vision image output is taken as information while a robotics perception system translates the image output from information into actions for taking decisions and actions in real world environment. Therefore, perception plays vital role for the success of goal-driven robotic system. However, despite this difference, robot perception incorporates the techniques from computer vision, and it is particularly evolving with the recent development in deep learning networks.

In real-world applications, endowing a robot with human like-perception for navigation is a challenging task that enables the robot to recognize scene and object when navigating through a dynamic complex environment and building a 3D map by observing the surrounding. Therefore, regardless of selected navigation system, object identification and place recognition play a vital role for environment representation and modeling.

### A. Object Recognition

Reliable object recognition is an important and early step for a mobile robot to achieve its goal. Real time object recognition systems work in two stages: Offline and online. Offline stage aims at reducing the execution time without affecting system efficiency. Image pre-processing, feature extraction, segmentation and training processes are performed in this stage. Online stage runs the process in real time to ensure the high-level interaction between robot and its surrounding environment. Image retrieval, classification, object detection and recognition are the examples of few processes that are carried out at this stage.

A key issue in this context is the interaction with object of different shapes and sizes. Despite significant achievements and advent of digital camera, accurate object detection and recognition is still a challenging task when real-world environment is considered. The reasons for this difficulty are occlusions, complex object shapes, variations in geometric and photometric pose, noise and illumination changes.

Early efforts [Zou19] to handle this issue are based on template matching. Later approaches include statistical classifiers including SVM, Adaboost and neural networks. On the other hand, computationally simple and efficient approaches based on local features such as scale-invariant descripts (e.g. SURF, SIFT), haarlike features also exist. However, the limitations of these methods include accuracy that depends on number of features that describe an image, segmentation that becomes highly complex in real world scenarios and not robust to relatively large affine transformations. In literature, its alternative is to use Object Action Complexes (OACs) [Petrick08] that combines the action, object and learning process to deal with the representational difficulties in diverse areas.

The perception-action relationship based on cognitive understanding has been explored in [Yan14] by linking both tasks through a memory component. In these studies, perception system uses three sensor modalities: vision, audio and touch and their data are passed to the memory module for generating the motor control signals and action unit translate them into robot responses. This intermediate process acts as robot's brain for improving the recognition task when mobile robot navigates in unknown environment. The study of attention based cognitive architecture in [Palomino16] uses the reasoning as a bound between perception and action. The core of this work is selection of active task based on the context data and accomplishment of task depends on the presence of specific element in the scene. However, object-based visual attention system still requires considerable efforts to accurately detect and categorize different objects. A recent study [Ye17] presents a vision system for object detection and recognition from a visual input in real time by computing motion, color, motion and shape cues and combining them in a probabilistic manner for assistive robots.

However, despite the vast analysis of existing perceptual systems for autonomous mobile robots, semantic recognition system remains to be addressed for robust object recognition in real-world scenario.

### B. Place Recognition

Visual place recognition becomes very challenging when real-world scenario is concerned. Therefore, visual place recognition algorithms must endow the autonomous mobile robot to robustly handle the variation in visual environment that occur due to dynamic, geographical and categorical changes [Martinez17]. The visual appearance of places varies due to illumination changes (day and night), moving the furniture or different objects from one place to another. The same place (room or corridor) might look different in different viewpoints, despite sharing some common visual features. Humans can recognize a room (office or kitchen) because of their ability to build categorical models of places. However, it is difficult for the robot to recognize the rooms based on their distinctive features and categories.

Literature review [Ullah08] shows that contextual understanding of the place is very important for autonomous mobile robot to effectively perform its task. A mobile robot can effectively interact with its environment if it recognized the place and have a functional understanding of area

## 2.4 Semantic Descriptor

There have been few empirical investigations in recognizing the objects that have semantic similarities in their shapes. A recent study [Tasse16], address this challenge and computes the semantic similarities between shapes,

images and depth maps using semantic based descriptors. The central idea is to combine labeled 3D shapes with semantic information in their labels for generating semantic-based 3D shape descriptor. An early study [Zen12], uses enhanced semantic descriptors for complex video scene understanding by embedding semantic information in the visual words. Recent developments in robot localization and mapping approaches have heightened the need to use semantic descriptors for robot localization and mapping. A seminal study in [Panphattarasap18], uses 4-bit binary semantic descriptor (BSD) for robot localization in 2-D map and performs semantic matching. The semantic features such as gap between buildings and road junctions are detected using CNN in urban environment. The purpose of BSD is to endow the robot with ability akin to human map reading.

### 3 Framework

Our proposed framework adds semantic techniques to SLAM to cope with the challenges in dynamic environment by providing the robot advanced perception that is closer to human vision and improving the world understanding capabilities of the robot for carrying out high-level navigation task in complex unstructured environments. Our framework provides a closer representation with global environment by defining the Triplet Ontological Semantic Model (TOSM) in which relations between the concepts are described for explaining semantic interoperability of environment.

#### 3.1 TOSM: Triplet Ontological Semantic Model

We accelerate the implementation of cognitive system in autonomous mobile robot by developing Triplet Ontological Semantic Model (TOSM) which is based on cognitive process of human perception and brain GPS model from neuroscience research and physiology. The main characteristics of TOSM are:

- To endow the robot with semantic mapping of environment based on cognitive architecture modeling
- To define the relations between domain concepts (knowledge), their attributes (properties) with high-level of abstraction and rules to reason based on the task and the environment
- To model the sensory information for performing the task planning

Our TOSM approach consisting of three major components for effective representation of domain knowledge and information retrieval in indoor environment is shown in Fig 1. Unique characteristics of these three components represent relationship information with different objects that have spatial and non-spatial properties for performing a specific task in overall robotic environment. The spatial properties represent the concepts of position, shape and size of the objects in robotic environment while the non-spatial properties determine the object category. We describe complete domain knowledge using spatial representation of the objects. Our proposed TOSM approach endows the robot to semantically map the objects and their positions in unexplored environment by defining explicit, implicit and symbolic models, shown in Figure Figure 1

##### A. Explicit Model

Explicit model specifies the spatial representation of the entities such as shape of an object and its position in the domain (global environment) by extracting all the geometrical features of that object and retrieving its physical information from sensors.

##### B. Implicit Model

Implicit model describes the behavior of the robot and series of actions such as robot navigation to perform a task. This spatial representation also defines the intrinsic relations between the entities, gives the semantic interpretation of environment which cannot be obtained using sensors and processes the fuzzy information to provide the effective interaction of the mobile robot with its surrounding along with planning capabilities. Introducing this model in our framework also enables the robot to take high-level decision by understanding the semantic concepts that constitute task success, such as it allows the robot to interpret the semantics of automatic door by understanding its salient events that auto-door opens and closes automatically, on sensing the approach of a person.

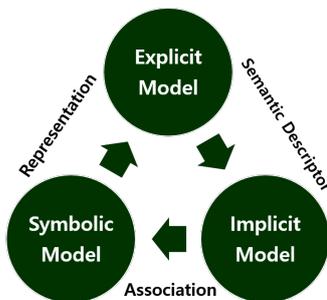


Figure 1: Triplet Ontological Semantic Model (TOSM)

### C. Symbolic Model

We use symbolic model to encode the domain knowledge for describing semantic descriptions, sequence of actions and complex capabilities of our environment in language-oriented way. Robot uses this knowledge through relations that are represented by the links between existing entities. Based on the integrated components of implicit, explicit and symbolic models, the TOSM approach coexist in SLAM allows the robots to perceive, learn, understand and interact with the surroundings based on geometric and semantic information.

### 3.2 TOSM on demand database

We design robot-mounted on-demand database to construct semantic model of the environment for providing the robot a semantic mapping and perception closer to human cognitive skills using TOSM. Our TOSM on-demand database approach has three main practical advantages:

- It eliminates the demand to store several different maps
- It generates the maps only when they are required for a robot to perform the assigned task in global dynamic environment
- It enriches the database semantically by adding conceptual meaning to data and relationships

We store, environmental and behavioral information together with robot knowledge and map data in on-demand database. Robot uses cloud database to plan the behavioral actions and on-demand database to builds a dynamic driving map according Figure 1: Triplet Ontological Semantic Model (TOSM) to assigned task in operating environment. If the robot needs additional information to download from network or cloud database for performing a specific task, this information is also merged with the robot’s current knowledge and on-demand database is concurrently updated. The on-demand database of environment based on TOSM describes the semantics of the domain with the set of relations. We have developed it using the protégé tool to explicitly represent the class hierarchy for each individual. Individuals, also called instances, are defined to represent a specific object in a class. For instance, automatic door is an individual of ‘Door’ class, as shown in Figure 2(a). We describe our ontological model by creating individuals (instances) in corresponding classes, connecting them with typed literals and defining relationships between objects of different classes. TOSM for on-demand database is composed of three main components: classes, object properties and data properties.

#### A. Classes

We use classes to describe the concepts using collections or types of objects that share common properties in indoor environment. Our ontological model consists of five classes: Map, MathematicalStructure, Time, Behavior and EnvironmentElement. Each class represents an abstract group of objects that belongs to the specific class. TOSM allows the classes to have single inheritance (one parent) and multiple inheritance. For example, subclasses of object, Occupant, Robot and Place in EnvironmentElement class have single inheritance while AutomaticDoor class has multiple inheritance. Thus, all the properties of parents’ classes (Door, Object and EnvironmentElement) are inherited by child class (AutomaticDoor). TOSM uses subclasses to represent the concepts more specifically than super classes. Figure 2(a) also shows that we have developed our class hierarchy with the systematic top-down view of domain in which we define the most general concepts of an entity in high-level (superclass) and more specific concepts in low-level (subclass).

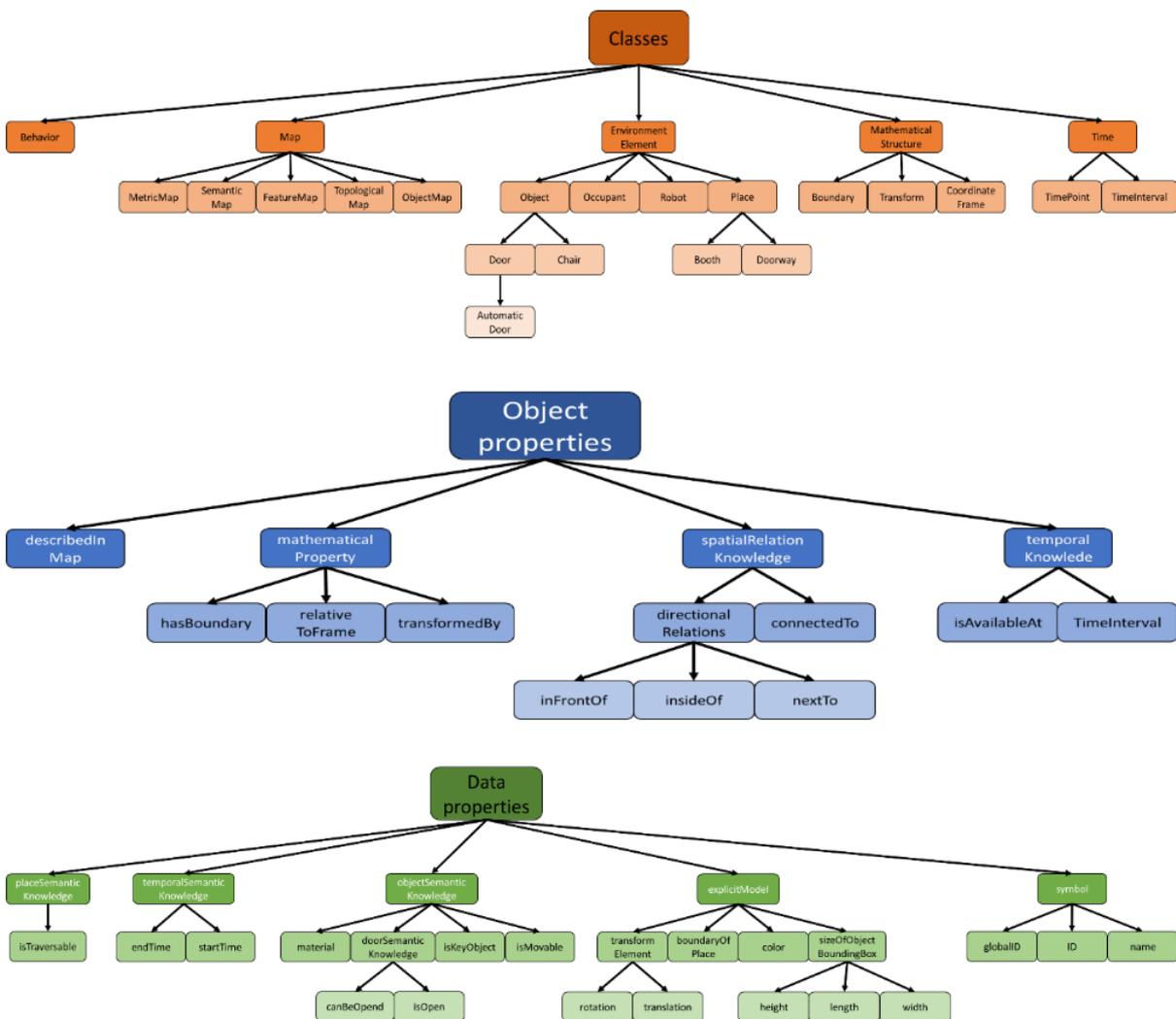


Figure 2: TOSM Properties for on demand database. (a) Class Properties; (b) Object Properties; (c) Data Properties

## B. Object Properties

These properties explain the relationship between the classes based on their instances. The category of object and set of properties determine the type of relationship between them. Figure 3 shows the expression of 3D geometric relation between two classes: “Room1 hasBoundary Boundary1” in which an object property “hasBoundary” links the individual “boundary1” of MathematicalStructure class to the individual “room1” of “EnvironmentElement” class. This geometric relation is inferred from visual perception and semantic map.

We divide the object properties into describedInMap, mathematicalProperty, spatialRelationKnowledge and temporalK. Figure 2(b) shows that mathematicalProperty includes hasBoundary, relativeToFrame and transformedBy, whereas spatialRelationKnowledge includes connectedTo and directionalRelations which is divided into inFrontOf, insideOf and nextTo. Finally, temporalKnowledge include isAvailableAt and TimeInterval properties

## C. Data Properties

These properties specify object parameters or typed literal, also called datatype (string, int, float). We retrieve the individuals by connecting them with the specified literal values using placeSemanticKnowledge, temporalSemanticKnowledge, objectSemanticKnowledge, explicitModel and symbol that are defined as data properties in

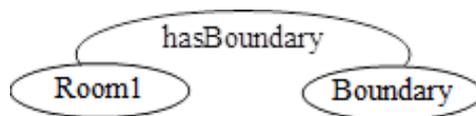


Figure 3: Geometric Relation between Two Classes

our ontological model, as shown in Figure 2(c).

### 3.3 Semantic descriptor-based Learning and Recognition

Our proposed framework introduces real-time and near-perfect object detection and place recognition approaches to mimic human visual system using semantic descriptor-based learning. The overview of our recognition model inspired by human visual cortex and semantic descriptor is illustrated in Figure 4.

When autonomous mobile robot explores complex indoor environment to perform a task, the perception module recognizes the objects and places by extracting data from sensors and retrieving from on-demand TOSM database. It continuously updates symbolic state of the task based on semantic information of newly obtained data from sensors and adds the implicit data about novel objects and places by identifying their classes in knowledge base.

Our framework allows open-ended learning that enables the robot to adapt to new environment by acquiring the knowledge in incremental fashion and accumulating conceptualization of new object categories. Apart from extensive training data for learning, a robot might always be confronted with an unknown objects and places in operating environment. Our framework handles this issue by processing visual information continuously and performs learning and recognition simultaneously. Our recognition model performs object detection and place recognition using convolutional neural network and semantic descriptor that is based on human perception system. The overview of our recognition model is described in Figure 4.

Our proposed recognition model consists of two stages: Training stage and Testing stage

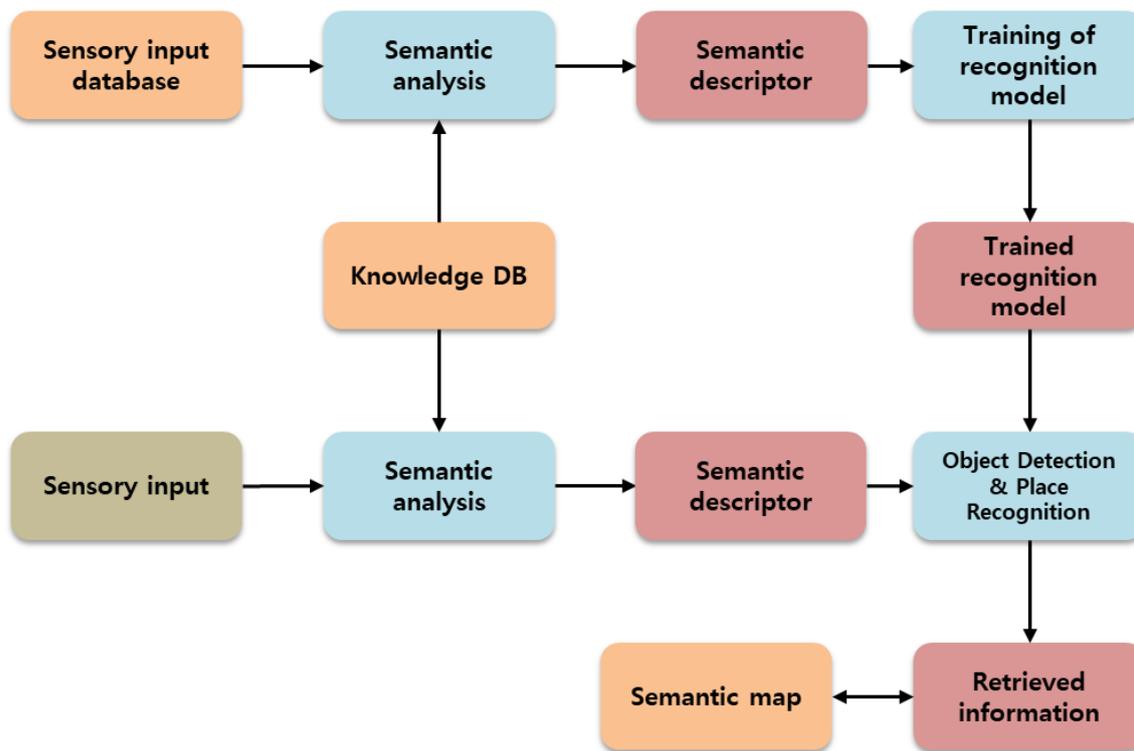


Figure 4: Semantic Descriptor-based Learning and Recognition Model

### A. Training Stage

At training stage, we use CNN to train the object detection and place recognition model using our own indoor dataset for making the prediction using sensory input data and on-demand database. This stage is composed of three major components: semantic analysis, semantic descriptor and training the recognition model.

We perform semantic analysis for explicit and implicit models to get the semantic information and characteristics of each object. Two major operations related to preprocessing of visual data and feature extraction are involved in this step. We perform preprocessing to improve the performance of recognition model by reducing the noise from data for better local and global feature extraction and detection. We extract semantic object features from processed visual data including both global features and local features. We get the overall properties of each object by extracting the global features (edges, corners and color) while salient regions by retrieving the local features.

For semantic analysis, we extract the geometric features such as edges, lines, corners and shape in conjunction with metric information related to size and pose estimation of object are extracted and integrated into explicit model of our framework as global features. We store object properties and relationship between them as sensory input data while actions of an object such as movability as information of object’s behavior in on-demand database.

The result of object analysis at semantic level is the extraction of semantic descriptions as per human perception. Thus, we reduce the semantic gap by combining the visual features extracted at low-level and information at high-level using semantic descriptor. We pass features vectors containing the geometric properties of the objects such as edges instead of the whole image to train our recognition model.

### B. Testing Stage

At this stage, we run our recognition model in real world by performing the semantic analysis on the visual data and passing the feature vector to run our trained CNN model for object and recognition. Computational simplicity and minimum storage requirements are the major motivating factors for us to pass the extracted feature vectors instead of whole image to the recognition model. It also endows the robot with the ability of human-like perception and semantic understanding of the environment.

## 4 Experiment

We perform the real-world experiments in conventional center to evaluate the performance of our proposed semantic SLAM framework and extract the information of the environment and objects. These evaluations are conducted on an Intel Core i7-4712MQ 2.30 GHz CPU, NVIDIA GeForce 840M GPU, and 12GB RAM. Our recognition module uses ZED camera to detect objects and places while we perform localization and mapping using the data obtained from laser range finder (2D sensor).

We use TOSM to represent semantic information by establishing the concepts and linking the conceptual and physical objects of the environment. Figure 5. Shows the model of our environment in which operating areas is highlighted in red color.



Figure 5: Experiential Real-world Environment

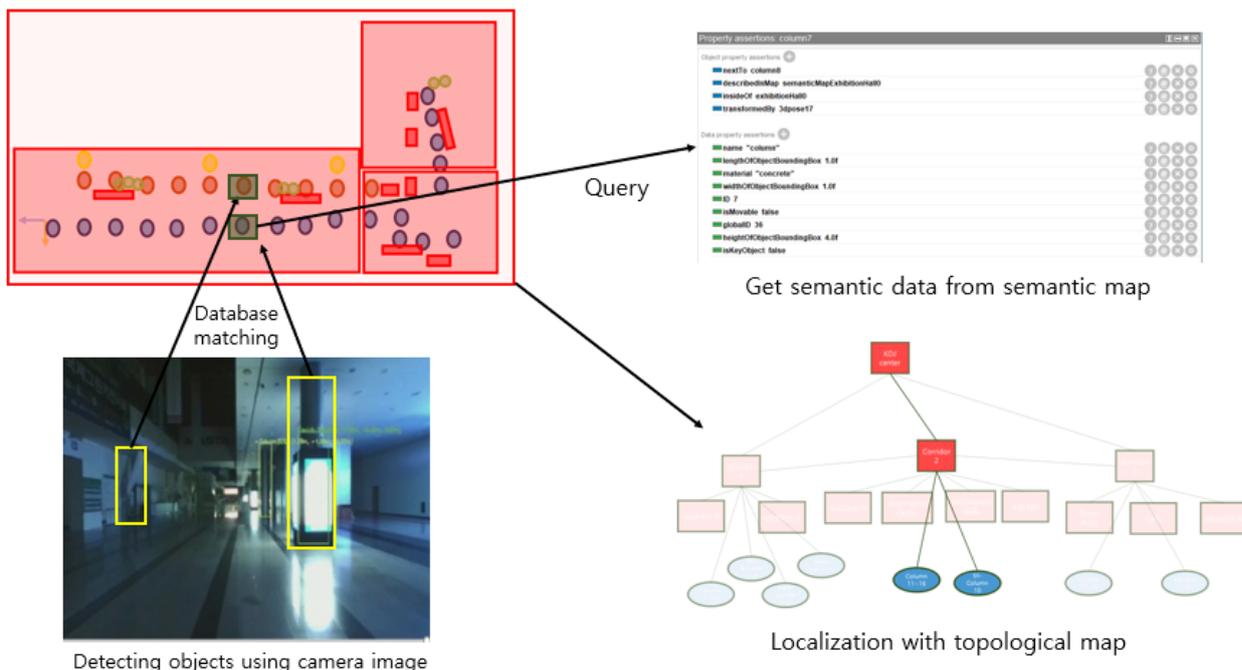


Figure 6: Experimental Result. (a) Recognized objects, (b) Semantic representation of Environment (c) Semantic database, (d) Topological map linking semantic environment with geometric information

Figure 6 shows all the steps involved in our real-world experiment. The robot localizes itself using topological map that endows the robot with spatial awareness. We build the semantic map by establishing semantic relationship between a place node in a topological map and its concepts. After that, we associate the objects that are recognized in a specific place with their topological nodes in the semantic map. Robot connects to the database that stores semantic information and properties of objects and places in order to match the relations. Figure 6(a) shows the objects recognized by our recognition model.

Our semantic map explains the structure of the environment at higher level that is closer to human-mapping system. Figure 6(b) shows the semantic representation of the environment, in which places are represented by rectangular boxes and objects are represented by circles. The blue circles indicate the columns, the orange circles show the vending machine along with the green circles that represent tri-columns while red boxes are places. Figure 6(c) shows the database that stores the ontology information for the robot mapping system and properties of the physical objects that are recognized when robot navigates in the environment. Our topological map shown in Figure 6(d) represents the environment by linking geometrical information and establishing the relations of semantic information to edges and nodes of relation graph. The proposed relation graph is focused on the environment mapping task and demonstrates semantic knowledge with the conceptual and spatial hierarchy. It represents the relationships between the information of corridor-1 containing elevators and columns along with the objects in coordior-2 and coordir-3 that the robot knows.

We extract semantic map of our environment model based on occupancy grid as shown in Figure Figure 7 and add semantic concepts such as corridors and spatial relations like connectivity between different objects in the environment.

## 5 Conclusion

In our semantic SLAM framework, we have presented the central idea to endow the mobile robot with intelligent behavior. It has introduced the biological vision-based perception system for object and place recognition using CNN and semantic descriptor. Furthermore, we have proposed human brain inspired semantic mapping system to modulate the robot's behavior when it navigates in the environment to perform a task. Moreover, our TOSM approach represents the knowledge about the elements in the map. The experimental results indicate the feasibility of our prof framework in real-world indoor environment. In the future we plan to investigate to build

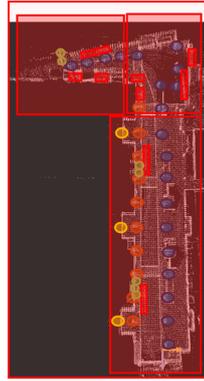


Figure 7: Semantic Map with Occupancy Grid Map

and updating a semantic map automatically without traditional maps and recognize objects and places using semantic map.

### Acknowledgement

This research was supported by Korea Evaluation Institute of Industrial Technology (KEIT) funded by the Ministry of Trade, Industry & Energy (MOTIE) (No. 1415162366 and No. 141562820)

### References

- [bohgl7] J. Bohg et al., Interactive perception: Leveraging action in perception and perception in action *IEEE Trans. Robot* – Volume. 33, no. 6, pp. 1273–1291, 2017.
- [Capobianco15] R. Capobianco, J. Serafin, J. Dichtl, G. Grisetti, L. Iocchi, and D. Nardi, A proposal for semantic map representation and evaluation, *2015 European Conference on Mobile Robots (ECMR) 2015 - Proc.*, pp. 1–6, 2015.
- [Pendleton17] S. Pendleton et al. “Perception, Planning, Control, and Coordination for Autonomous Vehicles,” *Machines* – Volume. 5, no. 1, p. 6, 2017.
- [Vasudevan08] Vasudevan, S. and Siegwart, R. Bayesian space conceptualization and place classification for semantic maps in mobile robotics *Robotics and Autonomous Systems* – Volume. 56, no. 6, pp. 522–537, 2008.
- [Burgard07] W. Burgard, P. Jensfelt, R. Triebel, Ó . Martínez Supervised semantic labeling of places using information extracted from sensor data *Robotics and Autonomous Systems* – Volume. 55, no. 5, pp. 391–402, 2007.
- [Nüchter08] A. Nüchter and J. Hertzberg. Towards semantic maps for mobile robots *Robotics and Autonomous Systems* – Volume. 56, no. 11, pp. 915–926, 2008.
- [Xie13] D. Xie, S. Todorovic, and S. C. Zhu. Inferring “dark matter” and “dark energy” from videos *IEEE International Conference on Computer Vision* – pp. 2224–2231, 2013
- [cadena16] Cadena, Cesar, et al. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age” *IEEE Trans. Robot.*, *IEEE Transactions on robotics* – Volume. 32, no. 6, pp. 1309–1332, 2016
- [Lowry16] S. Lowry et al. Visual Place Recognition: A Survey *IEEE Transactions on Robotics* – Volume. 32, no. 1, pp. 1–19, 2016
- [Sabourin10] C. Sabourin and K. Madani. Towards Human Inspired Semantic Slam *ICINCO 2010 - Proceedings of the 7th International Conference on Informatics in Control, Automation and Robotics* – Volume. 2, pp. 360–363, 2010

- [Pronobis12] A. Pronobis and P. Jensfelt. Large-scale semantic mapping and reasoning with heterogeneous modalities *IEEE International Conference on Robotics and Automation* – pp. 3515–3522, 2012.
- [Bowman17] S. L. Bowman and G. J. Pappas. Probabilistic Data Association for Semantic SLAM *IEEE International Conference on Robotics and Automation (ICRA)* – pp. 1722–1729, 2017.
- [Durand07] N. Durand et al. Ontology-based object recognition for remote sensing image interpretation *19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007)* – Vol. 1, pp. 472–479, 2007.
- [Ji12] Z. Ji et al. Towards automated task planning for service robots using semantic knowledge representation *IEEE 10th International Conference on Industrial Informatics* – pp. 1194–1201, 2012.
- [Riazuelo15] L. Riazuelo et al. RoboEarth Semantic Mapping: A Cloud Enabled Knowledge-Based Approach *IEEE Transactions on Automation Science and Engineering* – Volume. 12, no. 2, pp. 432–443, 2015.
- [Saigol15] Z. Saigol, M. Wang, B. Ridder, and D. M. Lane. The Benefits of Explicit Ontological Knowledge-Bases for Robotic Systems *Towards Autonomous Robotic Systems* – pp. 229–235, 2015.
- [Azevedo18] H. Azevedo, J. P. Ribeiro Belo, and R. A. F. Romero, OntPercept: A Perception Ontology for Robotic Systems *Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)* – pp. 469–475, 2018.
- [Lee18] S. Lee and I. Kim, A robotic context query-processing framework based on spatio-temporal context ontology *Sensors* – Volume. 18, no. 10, 2018.
- [Sünderhauf18] N. Sünderhauf et al. The limits and potentials of deep learning for robotics *The International Journal of Robotics Research* – Volume. 37, no. 4–5, pp. 405–420, 2018.
- [Zou19] Z. Zou, et al. Object Detection in 20 Years: A Survey *arXiv preprint arXiv:1905.05055* – pp. 1–39, 2019.
- [Petrick08] R. Petrick et al. Representation and Integration: Combining Robot Control, High-Level Planning, and Action Learning *Proceedings of the 6th international cognitive robotics workshop* – pp. 32–41, 2008.
- [Yan14] Yan, H., Ang, M.H. and Poo, A.N. A Survey on Perception Methods for Human-Robot Interaction in Social Robots *International Journal of Social Robotics* – Volume. 6, no. 1, pp. 85–119, 2014.
- [Palomino16] Palomino, A.J., Marfil, R., Bandera, J.P. and Bandera, A. A new cognitive architecture for bidirectional loop closing *Robot 2015: Second Iberian Robotics Conference* – Volume. 418, no. November, pp. 721–732, 2016.
- [Ye17] Ye, Chengxi, et al. What can i do around here? Deep functional scene understanding for cognitive robots *IEEE International Conference on Robotics and Automation (ICRA)* – pp. 4604–4611, 2017.
- [Martinez17] Martinez-Martin, E. and Del Pobil, A.P. Object detection and recognition for assistive robots: Experimentation and implementation *IEEE Robotics & Automation Magazine* – Volume. 24, no. 3, pp. 123–138, 2017.
- [Ullah08] Ullah, Muhammad Muneeb, et al. Towards robust place recognition for robot localization *IEEE International Conference on Robotics and Automation* – pp. 530–537, 2008.
- [Tasse16] Tasse, F.P. and Dodgson, N. Shape2vec: Semantic-based descriptors for 3D shapes, sketches and images *ACM Transactions on Graphics (TOG)* – Volume. 35, no. 6, pp.1–12, 2016.
- [Zen12] Zen, Gloria, et al. Enhanced semantic descriptors for functional scene categorization *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)* – pp. 1985–1988, 2012.
- [Panphattarasap18] Panphattarasap, P. and Calway, A. Automated Map Reading: Image Based Localisation in 2-D Maps Using Binary Semantic Descriptors *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* – pp. 6341–6348, 2018.