Ontology Learning of New Concepts Combining Textural Knowledge, Visual Analysis, and User Interaction

LILIANA VILLAMAR GÓMEZ and JUN MIURA, (Member, IEEE)
Department of Computer Science and Engineering, Toyohashi University of Technology, Toyohashi 441-8580, Japan
Corresponding author: Liliana Villamar Gómez (liliana@aisl.cs.tut.ac.jp)

This work was supported in part by the Japan Society for the Promotion of Sciences (JSPS) KAKENHI under Grant 17H01799.

ABSTRACT The continuous emergence of new technologies has contributed to the impending reality of service robots an upcoming reality. When interacting with humans, robots must adapt to changing environments. Hence, service robots at home need learning capabilities to acquire new knowledge and merge it with their own. In this study, we have developed a system for learning the ontologies of new concepts, combining textural knowledge, visual analysis, and user interaction. In this system, the robot is provided with an essential feature to adapt to the home environment. We focus on the learning of new ontological concepts oriented toward service robot applications. We propose combining textural knowledge, visual analysis, and user interaction to determine the correct placement of the new concepts in the ontology structure. We aim to enable the robot to extend its ontological knowledge as needed. We conducted a set of experiments to show the applicability of the presented method and the advantage of conceptualizing objects in ontological knowledge. The experiments consisted of two parts: concept learning experiments and experiments with an integrated robot system. In the former, the robot had to conceptualize a set of new objects in its ontological knowledge, and in the latter, the robot was asked to search and find the new objects learned.

INDEX TERMS Concept learning, ontology learning, robot learning, human–robot interaction.

I. INTRODUCTION
The continuous emergence of new technologies has contributed to the impending reality of service robots. Service robots are used in areas such as hospitality with great success in supporting customers and employees. A particular case is the service robots at home, which assist humans closely in performing household chores. These types of robots require standard capabilities such as navigation and vision, as well as personalized means of interaction according to the robot’s purpose. An essential part of developing such robots is knowledge management, which contains information regarding the scenes and robot; this knowledge is generally integrated beforehand. However, robots must adapt to the changing environments when interacting with people. Hence, service robots at home need learning capabilities to acquire new knowledge and merge it with their own.

The use of ontological knowledge to represent concepts has been proved to be beneficial in many robot applications. It gives a broad conceptualization of the meaning of objects and allows for the connection of related concepts, making knowledge extendable. Ontology learning has been achieved for different domains using linguistic and statistical techniques and inductive logic programming. These techniques include data preprocessing, relation extraction, and term extraction [1].

Expanding robot knowledge had been achieved using multiple techniques, including the use of language and perceptual information [2]–[4]. However, a broad and a clear conceptualization of objects is not contemplated. Different systems that include a hierarchy of concepts had been proposed [5], [6]. However, the robots’ autonomous learning is not considered.

In this study, we develop a system for ontology learning of new concepts combining textural knowledge, visual analysis, and user interaction. In this system, a robot is provided with an essential feature to adapt inside a home environment.
We focus on the learning of new ontological concepts oriented to service robot applications. We propose combining textural knowledge, visual analysis, and user interaction to determine the correct placement of these new concepts in the ontology structure. We aim to enable the robot to extend its ontological knowledge as needed.

The contributions of this work are as follows: (a) textural knowledge acquisition for word meaning identification and image data collection, (b) image visual analysis for concept description selection, (c) user interaction to support meaning selection, and (d) ontological knowledge update with the conceptualization of new objects.

Our previous work [7] showed an early version of this system, which used a basic visual analysis. It demonstrated the importance of selecting the proper meaning of a new concept during the learning process. In this work, we have included the utilization of neural networks in the visual analysis process and human–robot interaction to support meaning selection. In addition, we conducted experiments, including experiments with a complete service robot application scenario.

The rest of this study is organized as follows. Section II describes related work. Section III explains the problem definition and the environment layout. Section III-A briefly describes ontology learning components included in this system. Section IV and V explain word meaning identification and visual analysis processes, respectively. Section VI describes concept description selection. Section VII explains the ontology update process. Section VIII shows the experiments in learning a new concept. The study ends with the conclusions and discussion in Section IX.

II. RELATED WORK

Ontology learning techniques have been explored in different research domains such as tourism [8], [9], medicine [10], [11], and computer science [12], [13]. In addition, researchers have developed tools for ontology development tasks, such as the ROBOT tool [14]. This tool helps to check logical errors and standard quality control; it eases the creation, maintenance, and release of ontologies, currently working only with Open Biomedical Ontologies [15].

In the field of robotics, an attempt to create an ontology standard for autonomous robotics was presented in [5]. The authors demonstrated some advantages of using ontologies for a system, such as semantic interoperability. The knowledge representation focuses on notions such as behavior, function, goal, and task. Although the ontology includes object manipulation behaviors, it is unclear how objects are described. In addition, it is not mentioned if robots can extend their knowledge autonomously.

An approach to expand the knowledge of a conversational agent during runtime was presented in [2]. It presents a technique for automatic knowledge extraction, and it was tested with a social humanoid robot in a residential care home setup. The insertion of concepts is accomplished using verbal interaction, and it includes cultural knowledge about the user. Before the robot inserts a new concept, it asks whether the new concept belongs to a previously mapped class. However, it will continue asking in the same way for each class of ontology in a descending form, which might lead to a rather prolonged confirmation process if the mapped class is too general. Furthermore, the correct insertion of concepts depends on the ability of a user to describe the new concept so that the system can find a matching or related class; this can limit meaning conceptualization.

A different approach for learning concepts is found in [3], where an integrated cognitive architecture was realized to learn concepts, actions, and language. The authors integrated multiple probabilistic generative models to accomplish this. Experiments proved that a robot could learn based on its experiences while interacting with objects and receiving feedback from a user. However, they do not consider any hierarchy or relation between concepts related to objects learned by the robot. In addition, the model does not consider reasoning on the newly acquired concepts.

Learning word meanings has been achieved using statistical models such as probabilistic latent semantic analysis and latent Dirichlet allocation [16]. The approach, further extended in [4], used multimodal information such as visual, auditory, and haptic information, and it can learn object concepts incrementally. However, the categorization of objects is based on their visual features, so the semantic relationships between the object concepts are not present.

Another approach for learning visual concepts was proposed in [6]. It collects training videos and uses human annotation for data collection. The vision system learns concepts that are used to extract related concepts with ontology knowledge. The system searches for more videos using those extracted concepts and repeats the process. It shows the importance of having a concept hierarchy. However, it is not a fully autonomous learning process since it relies on human annotation data.

In [17], an incremental object learning system was described. The system uses a few sample images to learn new objects and their categories visually. However, the category hierarchy of concepts is not considered in the learning process.

In the present work, we aim to allow a robot to increment its ontological knowledge of object concepts. The conceptualization of a new object is accomplished using textural knowledge. In addition, selecting the accurate meaning describing a new concept is supported by visual analysis and human–robot interactions. The textural knowledge contains information about the semantic relations of concepts, providing accurate descriptions of new objects. Furthermore, the representation of object concepts inside the ontological knowledge gives the advantage of reasoning on concepts and a more extended language description.
II. PROBLEM DEFINITION AND OUR APPROACH
A robot, with general ontological knowledge, is required to learn a new object, which entails learning the new word concept to correctly create the new instance and the corresponding classes inside its predefined knowledge. Although the robot could use methods, such as using online resources or actively asking a user about the new object to extend its knowledge, we want the robot to learn new concepts as needed with the little burden on the user as possible.

For this purpose, we want to achieve the learning of a new concept using three methods. First, (1) using textural knowledge to identify the possible meaning of a new word; second, (2) performing meaning selection by analyzing the visual characteristics of related concepts against the new object; third, (3) creating user interactions to support meaning selection.

The proposed scenario to demonstrate the usability and feasibility of learning new concepts through textural knowledge and visual analysis is shown in Fig. 1. A service robot with vision capabilities and ontological knowledge is requested to learn a new object. First, the robot receives the name and the image of the new object. Then, the new object must be conceptualized to include it in the current knowledge.

IV. WORD MEANING IDENTIFICATION
A. TEXTURAL KNOWLEDGE ACQUISITION
An important factor for inserting a new concept in the ontology hierarchy is determining the correct correlated concepts and the corresponding categories. The ontological knowledge contains a collection of concepts of objects associated with each other by properties such as the “subclass of” property. It also contains instances of real objects belonging to some of those class concepts. To extend this knowledge, the robot needs to acquire textural knowledge containing semantic relations associated with the new concept. Therefore, we decided to employ online processing resources to acquire the semantic relations required.

Textural knowledge is based on a well-known lexical database WordNet [18]. To access the WordNet database, we use the general architecture for text engineering GATE [19], which can add WordNet as a processing resource. First, the robot queries GATE with the new word concept to retrieve the corresponding collection of noun POS tag senses only. Each sense has a possible meaning for the new word concept as a noun and its semantic relations with associated concepts. Then, a list of hypernyms for each sense is obtained. This list represents the superclasses of the new word concept according to each sense that will be used to create the new concept in the ontology.

An example of a query for the word “pen” is shown in Fig. 3. The robot receives five different senses for the same word: writing implement, enclosed area, portable enclosure, correctional institution, and female swan. In Fig. 3 (bottom), the retrieved hypernyms for the first two senses are shown. After the robot obtains the list of senses and their semantic relations, it must identify which sense best describes the new word concept to add the corresponding semantic relation in the ontology.

B. SEMANTIC RELATION SELECTION
The selection of the correct semantic relation is a crucial step for ontology learning since this will establish the connection of a new concept with existing ones. In addition, new concepts that are correctly associated will allow the ontology to make inferences over them, such as deducing the possibly inherited characteristics from potential superclasses.

A. ONTOLOGY LEARNING COMPONENTS
Our approach for ontology learning of new concepts consists of four modules. These modules are responsible for (1) textural knowledge acquisition and semantic relation extraction, (2) image collection and analysis, (3) concept description selection, and (4) ontology updating. The modules are connected consecutively until finalizing with the object conceptualization process. These components enable the robot to learn new objects by conceptualizing them. The robot only needs the name of the new object and its image to start this process. It can also verbally interact with a user to obtain confirmation before conceptualizing a new object.

An overview of the ontology learning components is shown in Fig. 2. The concept learning process starts when a user shows the robot an image of the new object and names it; this information is sent to the Word Meaning Identification module. It acquires textural knowledge about the new object to extract its semantic relations as a concept. Then, all senses, which refer to the possible meanings of the new word concept, are sent, along with their semantic relations, to the Visual Analysis module.

The Visual Analysis module performs online image queries using the hypernyms contained in the semantic relations received. Next, it analyzes the images downloaded to find features similar to the image of the new object, and a similarity score is assigned to each sense. Senses with their respective similarity scores are sent to the Concept Description Selection module. In this module, the similarity scores are examined to choose a sense that best describes the new object concept to learn. When the robot finds more than one high similarity score, it resorts to interacting with the user to confirm the correct concept description. Once the robot knows which concept description best represents the new object, it sends it to the Knowledge Management module to update the ontology.

In the following subsections, we explain the word meaning identification process, the visual analysis performed, and the ontology update with new concepts.
The new semantic relation to be added to the ontological knowledge must comply with the following guidelines:

- The sense of the chosen semantic relation must correspond to the correct meaning of the new word concept.
- The new semantic relation will be linked to the closest or an equivalent concept in the ontology hierarchy.
- The concept classes of the new word concept will be added sequentially following the semantic relation.
- The insertion of a new class will stop when the new class already exists in the ontology. Otherwise, a maximum of four new classes will be created as the higher the concept in the hierarchy is, the more general it becomes.

Semantic relations acquired from WordNet might contain general concepts that can be found in the ontological knowledge, such as an artifact or physical entity (Fig. 3 bottom). However, this does not ensure that those concepts accurately describe the new concept to be learned. Analyzing the visual characteristics of the new object and user interaction aid in determining the correct semantic relation.
V. VISUAL ANALYSIS

We assume that the robot cannot recognize the new object with the current object identification module in the concept learning scenario. Hence, it is impossible to know the name or category of the object that could assert the position for the new semantic relation in the ontological knowledge.

Therefore, we propose supporting the selection of semantic relations by analyzing the visual characteristics of the new object and comparing it with objects belonging to the possible semantic fields of the new object based on its name. This visual analysis process aims to find the correct semantic field of the new object by studying the similarities with other objects from a set of potential semantic relations. The process starts with an online image query to collect images of hypernyms included in each potential semantic relation, followed by feature analysis, as explained in the following subsections.

A. ONLINE IMAGE QUERY

In the first part of the visual analysis process, the robot makes an online image query to collect image samples of object categories associated semantically. Potential categories are chosen from the list of potential semantic relations obtained previously in the word meaning identification process.

A query is formed using the new word concept and a hypernym, e.g., for the first sense of the concept washer, whose first three hypernyms are worker, person, and organism, the first query generated would be “washer worker.” This method of formulating queries helps the online image query return good results according to the expected meaning. For example, Fig. 4 shows image results for querying the hypernym of the second sense of the concept washer in different ways. According to WordNet, the second sense of the word washer refers to a “seal consisting of a flat disk placed to prevent leakage.” The difference in image results depends on the query. In the case of (a), (b), and (c), when the query contains only a hypernym, image results do not fully represent the meaning required. This situation can confuse the robot’s understanding since objects in the resulting images can be rather general and variable, not necessarily representing the expected meaning.

In contrast, adding the word washer to the query makes it more specific, and resulting images show objects closer to the meaning required (Figs. 4 (d-f)). Therefore, the robot can have a better approximation of objects belonging to the required sense.

Following the query generation method, the system creates online image queries for the first three hypernyms per sense to download sample images (Fig. 5). Thus, these images symbolize examples of objects from each semantic relation. Fig. 5 depicts the hypernyms of three senses corresponding to the concept washer and examples of the online image query results for the second and third senses.

The importance of correctly identifying the sense that best describes the new object to learn cannot be overstated. Choosing the wrong sense could result in a completely different meaning being assigned to the new object, the incorrect concepts, and an incorrect hierarchy being created in the ontological knowledge. Once the online image query collects sample images of the potential semantic relations for the new object, the next step is to determine the correct meaning for the new object by analyzing image features.

B. IMAGE ANALYSIS

The second part of the visual analysis process consists of comparing the new object’s image with the potential semantic relation data collection. With this, the robot is expected to find a group of images belonging to only one sense similar to the new object image.

We propose using a pretrained artificial neural network to extract image features of the new object and the downloaded collection of images. Then, it is possible to calculate a similarity value between them using the extracted features.

In this study, we use a deep convolutional neural network (CNN), namely, ResNet-152 version 2 architecture [20]. The CNN is pretrained on the ImageNet dataset [21]. First, we removed the network’s last layer to extract a 2048-dimensional feature per image from the last fully connected layer. Next, the features were computed for the new object image and the downloaded images of all the potential semantic relations.

Subsequently, the cosine similarity is computed between the features of the new object image and those of each image from the semantic relations. Then, the average similarity for each semantic relation, which represents a sense, is calculated. Hence, a similarity value is assigned to each sense (Fig. 6).

VI. CONCEPT DESCRIPTION SELECTION

Before the robot can add the new concept to its ontological knowledge, the final step is deciding which concept
description best represents the new object shown. Thus, to recap, the first part of the concept learning process is the word meaning identification, which acquires the meanings and the semantic fields of the new word concept. The second part is image analysis, which collects image samples for each semantic field and finds similarities between them and the new object to learn. Finally, the robot needs to choose the best concept description for the new object.
A. ISSUES IN CONCEPT DESCRIPTION SELECTION

The selection of the concept description is based on the results of the image analysis process, which outputs a list of similarity values of the new object for each sense. Ideally, the sense with the highest similarity score would be the best to describe the new object concept. However, the similarity is significantly affected by the variation of results during the online image query, as shown in Fig. 5.

There are some situations when image results may contain unrelated images, even with the descriptive query. That is the case of Fig. 7, where querying “pitcher containerful” retrieves mixed results. Therefore, it is necessary to emphasize that having additional images for each semantic relation contributes to a more accurate differentiation.

The senses of a word can refer to multiple different meanings for the same word concept. Hence, the robot needs to be sure that the correct concepts will be added to its ontological knowledge. There are three main cases to consider regarding the results of the similarities of images:

1) The sense with the highest similarity score is the correct one, and no other sense has a close similarity score.
2) The sense with the highest similarity score is the correct one, and a second sense has a close similarity score.
3) The sense with the highest similarity score is the incorrect one.

As previously mentioned, the first is the ideal case since the new object would be adequately conceptualized. However, for the second and third cases, an additional method is required to confirm that the correct sense is being selected. Therefore, we propose assisting in selecting the concept description using human–robot interaction, as explained in the following subsection.

B. USER INTERACTION FOR CONCEPT DESCRIPTION SELECTION

In this learning concept process, the robot is in a situation where it requires the user’s approval for the series of concepts that it is about to learn. Furthermore, as discussed in the previous chapter, human–robot interaction can be considerably helpful in a service robot environment.

We propose using human–robot interaction to provide final assistance in the concept description selection process, which has the following benefits:

- it helps the robot confirm that it is conceptualizing the new object concept correctly,
- it helps the robot decide which definition of the new object is accurate when the image analysis results are not confident,
- it gives assurance to the user that the robot is learning the new object correctly, and
- teaching a new object to the robot would be more interactive for the user.

We previously discussed the main scenarios that the robot may encounter after obtaining image similarities: when the
The highest score is either for the correct or incorrect sense, and when the second-highest score is close to the first. Depending on which of these scenarios the robot encounters, it interacts with the user throughout the concept learning process.

The robot creates three types of dialog sentences for (1) requesting the image and the name of the new object, (2) reporting that it saved the new object and a one-word description of it, and (3) allowing the user to choose between two options of concepts that possibly define the new object. While the first and second dialog sentences are always used in the concept learning to start and finish the process, the third dialog sentence is used when the first two highest similarity scores are very close. Hence, the robot can ask which one it should save.

The user interaction helps the robot with the final selection of the concept description. It is worth mentioning that limited interaction is preferable since spending much time teaching one object could be exhausting for the user.

VII. ONTOLOGY UPDATE
The final step in the concept learning process is to update the ontological knowledge with information about the new object. At this point, the correct semantic relation has been chosen according to the guidelines explained in subsection IV-B.

The ontology update process starts by creating a new instance of the object. Then, the first concept class created corresponds to the exact name of the object that is being taught; this means that if the new object is a “pen,” the first concept class would be Pen. Next, the following three hypernyms of the semantic relation will be added sequentially based on the hierarchy.

Finally, the creation of new classes stops when (1) the concept already exists in the ontological knowledge or when (2) four class concepts have been created. Figs. 8 and 9 show an example of these two cases, respectively.

One problem arises in the second case when four classes are created without finding an existing class. In this case, the four classes will not have any precedent class.
and would be forcefully linked to the root ontology concept. A new concept class linked to the root ontology will not connect with any other concept unless explicitly specified.

The robot has to conceptualize new tangible objects available in the current environment in the proposed concept learning scenario. Therefore, the previously mentioned problem is overcome by linking the last class created to the HumanScaleObject class, which best describes the possible objects the robot can learn (Fig. 9).

VIII. EXPERIMENTS IN LEARNING A NEW CONCEPT

We conducted a set of experiments to show the applicability of the presented method for ontology learning of new concepts. Some essential factors are considered in these experiments:

- Ontological knowledge enables the robot to progressively and accurately expand its knowledge. It also allows the utilization of more linguistic variations for the same referent.
- Textural knowledge contributes to the understanding of the meanings of a new concept.
- Visual analysis significantly supports the robot in the selection of the meaning of the new object concept.
- Human–robot interaction creates a more natural learning process and assists the robot in the correct conceptualization of objects.

A. EXPERIMENT SETUP

The experiments consisted of two parts: concept learning experiments and experiments with an integrated robot system. In the concept learning experiments (Section VIII-B), the robot was asked to learn new objects as follows (Fig. 10):

1) The user showed an image of the new object and named it.
2) The robot conceptualized the new object using textural knowledge and visual analysis.
3) In case more interaction with the user was required, the robot created it.

The concept learning experiments were divided into two tasks. In the first task, the robot learned one object, and the results were compared with a baseline method. In the second task, the robot learned eight different objects with specific characteristics to challenge the robot.

After the robot had finished learning the new objects of the concept learning experiment, we conducted the integrated system experiments (Section VIII-C), where the robot was asked to search and find the new objects learned.

B. CONCEPT LEARNING EXPERIMENT

The first experiment consisted of the robot learning only one object concept, and it did not include interaction with the user. This experiment is a continuation of experiments performed in [7]. This experiment aimed to show the importance of visual information to assist in selecting the semantic relation corresponding to the new object concept.

In the first experiment, we taught the robot the concept of a “pen.” The image and name of the new object were inputted to the robot. The robot started the learning process by acquiring the senses of the word “pen.” Then, it made an online image query of each of the hypernyms of each sense.
TABLE 1. Sample images of querying the hypernyms used in the experiments [7].

| Hypernym | Sample Images |
|----------|---------------|
| Writing Implement | ![Image](image1.png) |
| Enclosure | ![Image](image2.png) |
| Playpen | ![Image](image3.png) |
| Penitentiary | ![Image](image4.png) |
| Swan | ![Image](image5.png) |

With this, the robot created sets of sample images for each hypernym of the semantic relation. Sample images from the online image query are shown in Table 1.

As a baseline, we used color histograms in the visual analysis process for feature extraction. Hence, color histograms for all sets of images were created. Subsequently, the robot computed the cosine similarity between the histograms of each image searched online and the new object image. Finally, the average similarity was calculated for the set of images of each semantic relation.

The similarity score was assigned to each of the senses. The sense with the highest similarity score was selected to describe the new concept in the ontology. Table 2 shows the similarity scores calculated for each sense of the object “pen” using the baseline and our method.

TABLE 2. Similarity scores assigned to each sense of the object “pen” using the baseline and proposed methods.

| Word “pen” | Baseline | Our method |
|------------|----------|------------|
| Sense 1. Writing implement with a point from which ink flows. | 0.82 | 0.15 |
| Sense 2. Enclosure for confining livestock. | 0.54 | 0.11 |
| Sense 3. Portable enclosure in which babies may be left to play. | 0.70 | 0.13 |
| Sense 4. Correctional institution for those convicted of major crimes | 0.22 | 0.09 |
| Sense 5. Female swan. | 0.12 | 0.09 |

FIGURE 11. Objects used in the concept learning experiment. The names of the objects are: drill, durian, nail, pitcher, sponge, trunk, washer, and wrench.

The second experiment consisted of showing the robot eight different objects, which it must conceptualize. In this experiment, the robot used the proposed method for image analysis and user interaction. We demonstrated the challenges in a concept learning scenario involving objects and concepts.

In the visual analysis process, the robot used the pretrained network to extract image features. This time, the set of objects included the following types:

- The resulting images between each sense are visually similar.
- Images of the correct sense are significantly different.
- Images of the correct sense appear in the results of another sense.
- Very few images of the correct sense appear in the results.
- Images of the correct sense appear in the results of all senses.

Based on these experiments, we confirmed the importance of selecting the best sense describing the new object in the concept learning process. In addition, the correct placement of the new concepts in the ontology hierarchy is crucial for ensuring that the overall ontological knowledge is consistent.

The second experiment consisted of showing the robot eight different objects, which it must conceptualize. In this experiment, the robot used the proposed method for image analysis and user interaction. We demonstrated the challenges in a concept learning scenario involving objects and concepts.

In the visual analysis process, the robot used the pretrained network to extract image features. This time, the set of objects included the following types:
The image results are visually different for all senses.

- The correct semantic relation contains common concepts with the ontology.
- The correct semantic relation does not contain any common concepts with the ontology.

Fig. 11 shows the set of objects used for this second concept learning experiment. It includes a drill, durian, nail, pitcher, sponge, trunk, washer, and wrench. Each of these objects corresponds to at least one of the types of objects listed above.

During the experiment, the robot received the image of each object and its name sequentially. Next, it performed the word meaning identification process, acquiring the corresponding senses and semantic relations, as explained in Section IV. Then, it continued with the visual analysis process. In the online image query method, the robot created two datasets: a small dataset of 10 images and a large dataset of 50 images. Next, it extracted features using the pretrained network and computed the similarity scores for each semantic relation, as explained in Section V.

Table 3 illustrates the results of the similarity scores assigned to each sense of the set of objects for the small and large datasets. The senses marked in blue represent the true meaning of the new object concepts to learn. In addition, the highest similarity score for each sense per new object is marked in bold; this corresponds to the sense chosen by the robot.

Based on these results, we can see that the robot correctly chose the desired sense for all objects in both datasets. However, the similarity score is not as high as expected, and this is due to the variety of objects used in the previous experiments.

The chosen senses with a low similarity score belong to the object drill and the nail. The similarity score was significantly affected by the results of the online image query, where the images did not fully represent the expected meaning. In both cases, images similar to the new object appear only in the results of one hypernym of the chosen senses, as shown in Fig. 12.

In the case of the objects durian, sponge, and wrench, the highest two similarity scores are slightly close for the small dataset. These results were caused by images of the correct sense appearing in other senses’ results, increasing the score of the incorrect sense. However, we believe that having several samples for each sense according to its hypernyms adds more variations of objects helping in the semantic selection. This is evident in the results of sponge and wrench in the large dataset, where the two highest similarity scores are more distant than the scores for the same objects in the small dataset.

**Wrench**

| Sense 1. A sharp strain on muscles or ligaments. |
| Sense 2. A jerky pulling movement. |
| Sense 3. A hand tool that is used to hold or twist a nut or bolt. |

**Sponge** is another object in the small dataset with two close similarity values. However, the robot did not prompt a question with this concept due to the threshold value. Setting up a higher threshold might cause unnecessary interaction with the user. Therefore, an appropriate threshold calculation must be investigated in the future.
C. INTEGRATED SYSTEM EXPERIMENT

The last experiment consisted of the robot searching and finding the new objects in the current environment once it completed the concept learning process. The user challenged the robot skills to deduce the requested object and infer its location according to its knowledge.

In this last experiment, we demonstrated the advantage of having ontological knowledge when conceptualizing new objects, such as referring to the new object differently according to its newly connected classes and making inferences about the new objects using possibly inherited attributes.

The robot test included the following specifications to examine the integrated system after the concept learning effects:

- The objects are requested by their names or the names of their class.
- When the object class’s names are used, they are as follows: a directly stated class name, an inherited class name, and a possibly linked class.
- Object concepts that are expected to inherit attributes such as possible locations are requested.
- The integrated ontology-based knowledge management system with verbal interaction and concept learning was evaluated by its success in solving the object search experiment.

To provide the robot skills to complete the mentioned challenge, we joined the concept learning components proposed in this paper (see Fig. 2) and the ontology-based knowledge management system for home service robots presented in [22]. The system integration consists of command analysis, talking interaction, task planning, execution modules, and knowledge management connected to the concept learning components.

We conducted the experiments in the simulated environment used in [22], which includes a four-room house with static furniture and graspable objects commonly found in home settings. We placed boxes at different locations inside the living room, bedroom, lobby, and kitchen, as shown in Fig. 14. Each box had one of the images of the new objects that were used during the learning process.

The new objects were taught to the robot sequentially during the concept learning. Then, the robot was given different kinds of commands to find the newly acquired objects.

The conceptualization of the new concepts was tested using the following natural language commands, which evaluated the correct association of the concepts and the inheritance of features in the ontology:

1) Find the wrench.
2) Find the durian.
3) Find the sponge on the kitchen cabinet.
4) Find the nail in the bedroom.
5) Find the seal (washer).
6) Find the vessel (pitcher).
7) Find the luggage on the bed (trunk).
8) Find the tool in the lobby (drill).

| Object name | Sense number | Small dataset | Large dataset |
|-------------|--------------|---------------|---------------|
|              |              | Baseline | Our method | Baseline | Our method |
| drill       | Sense 1      | 0.58      | 0.19      | 0.53      | 0.15      |
|             | Sense 2      | 0.10      | 0.09      | 0.16      | 0.08      |
|             | Sense 3      | 0.43      | 0.11      | 0.50      | 0.10      |
|             | Sense 4      | 0.46      | 0.12      | 0.54      | 0.12      |
| durian      | Sense 1      | 0.09      | 0.29      | 0.01      | 0.26      |
|             | Sense 2      | 0.14      | 0.35      | 0.06      | 0.29      |
| nail        | Sense 1      | 0.73      | 0.10      | 0.36      | 0.10      |
|             | Sense 2      | 0.68      | 0.11      | 0.62      | 0.15      |
|             | Sense 3      | 0.62      | 0.10      | 0.43      | 0.10      |
| pitcher     | Sense 1      | 0.27      | 0.09      | 0.17      | 0.07      |
|             | Sense 2      | 0.66      | 0.30      | 0.58      | 0.27      |
|             | Sense 3      | 0.51      | 0.18      | 0.32      | 0.14      |
|             | Sense 4      | 0.44      | 0.11      | 0.22      | 0.09      |
|             | Sense 5      | 0.08      | 0.07      | 0.11      | 0.07      |
| sponge      | Sense 1      | 0.56      | 0.27      | 0.58      | 0.24      |
|             | Sense 2      | 0.41      | 0.21      | 0.38      | 0.17      |
|             | Sense 3      | 0.45      | 0.22      | 0.34      | 0.17      |
|             | Sense 4      | 0.03      | 0.15      | 0.13      | 0.14      |
| trunk       | Sense 1      | 0.00      | 0.09      | 0.00      | 0.09      |
|             | Sense 2      | 0.00      | 0.23      | 0.00      | 0.22      |
|             | Sense 3      | 0.00      | 0.12      | 0.00      | 0.12      |
|             | Sense 4      | 0.00      | 0.16      | 0.00      | 0.14      |
|             | Sense 5      | 0.01      | 0.10      | 0.01      | 0.09      |
| washer      | Sense 1      | 0.31      | 0.13      | 0.51      | 0.12      |
|             | Sense 2      | 0.63      | 0.27      | 0.54      | 0.25      |
|             | Sense 3      | 0.60      | 0.14      | 0.58      | 0.13      |
| wrench      | Sense 1      | 0.52      | 0.17      | 0.63      | 0.16      |
|             | Sense 2      | 0.77      | 0.24      | 0.72      | 0.19      |
|             | Sense 3      | 0.85      | 0.26      | 0.84      | 0.23      |
TABLE 4. Types of commands used for the integrated system experiments. They are divided by commands referring to the object by its name or class. The locations of the objects are divided into non-inherited and inherited if it is not given, and furniture and room if it is given. The number of instances available in the environment for those objects is listed.

| Command with an object called by | Location not given | Location given | Number of instances |
|-------------------------------|-------------------|----------------|-------------------|
|                               | Non-inherited     | Inherited      | Furniture         | Room   |
| Name:                         |                   |                |                   |       |
| 1. Find the wrench.          | O                 | O              | O                 |       |
| 2. Find the durian.          |                   |                |                   |       |
| 3. Find the sponge on the kitchen cabinet. | | | | |
| 4. Find the nail in the bedroom. | | | | |
| Class:                        |                   |                |                   |       |
| 5. Find the seal (washer).   | O                 | O              |                   |       |
| 6. Find the vessel (pitcher).|                   |                |                   |       |
| 7. Find the luggage on the bed (trunk). | | | | |
| 8. Find the tool in the lobby (drill). | | | | |

The first four commands request for the new objects by their names. The first two (“Find the wrench” and “Find the durian”) do not include a location to search for the object. The third command (“Find the sponge on the kitchen cabinet”) includes the name of the furniture. The fourth command (“Find the nail in the bedroom”) includes the room’s name to find the object.

The last four commands request for the new objects using the name of an upper class and two do not include a location to search for the objects. In the fifth command (“Find the seal”), the class name used refers to the object “washer,” and the sixth command (“Find the vessel”) refers to the object “pitcher.” The seventh command (“Find the luggage on the bed”) includes the furniture on which the object is located, and the class name refers to the object “trunk.” Finally, the eighth command (“Find the tool in the lobby”) includes the room’s name, and the class name refers to the “drill.”

In addition, the second and sixth commands (“Find the durian” and “Find the vessel”) include objects that are expected to inherit a location where are commonly found from its ontological knowledge. A particular case is considered in the eighth command (“Find the tool in the lobby”), where more than one object belonging to the required class exists in the environment. A summary of the characteristics of each command can be found in Table 4.

The robot completed all the commands using the integrated system successfully during the experiment. The robot used its ontological knowledge and interacted with the user when necessary. We analyzed the robot’s process in each command to evaluate the use of its ontological knowledge after the learning process. The results can be found in Table 5. The robot found object instances applicable to the description when the objects were requested either by their names or classes. Furthermore, the robot found a suitable instance of a location to search for the object when its ontological knowledge had an inherited location or when the command provided it.

A particular case was observed when the commands included any object belonging to the Tool class, which are “wrench,” “nail,” “washer,” and the “drill.” This type of class does not include an inherited location that could be inferred from the robot’s knowledge; therefore, the robot could not find a suitable location. However, since the integrated system includes human–robot interaction, the robot could ask the user for an object’s location. Another case occurred in the last command (“Find the tool in the lobby”), where the robot found more than one object that could belong to the Tool class, and it asked the user for the specific object.

We can conclude that the new objects learned by the robot were accurately conceptualized in the ontological knowledge, considering that the robot could find the possible instances of the requested objects. Furthermore, the new concepts learned for each new object were linked to the proper hierarchy since the robot found inherited properties such as the location.
TABLE 5. Experimental results using the integrated system. The robot found the instances of the requested objects using its knowledge or by asking the user. The locations were acquired using its knowledge or from the command, or by asking the user.

| Command                        | Object | Location       |
|--------------------------------|--------|----------------|
| 1. Find the wrench.            | O      | In knowledge   |
| 2. Find the durian.            | O      | In knowledge   |
| 3. Find the sponge on the kitchen cabinet. | O     | In knowledge/command |
| 4. Find the nail in the bedroom. | O      | In knowledge/command |
| 5. Find the seal (washer).     | O      | In knowledge   |
| 6. Find the vessel (pitcher).  | O      | O              |
| 7. Find the luggage on the bed (trunk). | O     | O              |
| 8. Find the tool in the lobby (drill). | O      | Succeeded      |

In cases, where the ontological knowledge insufficient, the system used human–robot interaction to acquire information, which is the expected behavior of a service robot.

IX. CONCLUSION AND FUTURE WORK

This study describes ontology learning of new concepts combining textural knowledge, visual analysis, and user interaction for service robot applications. We proposed analyzing the semantic relations and visual features of a new object concept to determine its correct conceptualization in an ontology. In addition, we employed human–robot interaction to assist the robot when necessary.

We tested the proposed concept learning method in a scenario, where the robot has to conceptualize new objects with only the image and the name of the object. In addition, we demonstrated some challenges to consider in concept learning scenarios. Future work could make some improvements, such as improving the input query formulation for the online image query process to obtain more accurate image results according to each sense. This improvement could be achieved by further analyzing concept meanings and the semantic relations of a concept since more information defining each sense can be obtained.

Furthermore, the interaction with users during the word meaning selection process could be improved by including more variations of questions formulated by the robot, such as inquiring about the new object and mentioning its possible applications. Improving the verbal interaction would be helpful too during the ontology update process, as there is a possibility of concept inconsistency. Ontology inconsistencies occurs when a concept is not adequately defined, including missing equivalent classes, disjoint classes, and other properties. Hence, a more extensive verbal interaction would be necessary.

REFERENCES

[1] M. N. Asim, M. Wasim, M. U. G. Khan, W. Mahmood, and H. M. Abbasi, “A survey of ontology learning techniques and applications,” Database, vol. 2018, Oct. 2018, Art. no. bay101, doi: 10.1093/database/bay101.

[2] L. Grassi, C. T. Recchiuto, and A. Sgorbissa, “Knowledge triggering, extraction and storage via human-robot verbal interaction,” CoRR, vol. abs/2104.11170, pp. 1–19, Apr. 2021. [Online]. Available: https://arxiv.org/abs/2104.11170

[3] K. Miyazawa, T. Horii, T. Aoki, and T. Nagai, “Integrated cognitive architecture for robot learning of action and language,” Frontiers Robot. AI, vol. 6, p. 131, Nov. 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/frobi.2019.00131

[4] T. Araki, T. Nakamura, T. Nagai, K. Funakoshi, M. Nakano, and N. Iwahashi, “Online object categorization using multimodal information autonomously acquired by a mobile robot,” Adv. Robot., vol. 26, no. 17, pp. 1995–2028, Dec. 2012, doi: 10.1080/10453866.2012.728693.

[5] J. I. Olszewski, M. Barreto, J. Bermejo-Alonso, J. Carbonera, A. Chibani, S. Fiorini, P. Goncalves, M. Habib, A. Khamis, A. Oliavera, E. P. de Freitas, E. Prestes, S. V. Ragavan, S. Redfield, R. Sanz, B. Spencer, and H. Li, “Ontology for autonomous robotics,” in Proc. 26th IEEE Int. Symp. Robot Hum. Interact. Commun. (RO-MAN), Aug./Sep. 2017, pp. 189–194.

[6] S. Arora, C. Cho, P. Fitzpatrick, and F. Scharffe, “Towards ontology driven learning of visual concept detectors,” CoRR, vol. abs/1605.09757, pp. 1–4, May 2016. [Online]. Available: http://arxiv.org/abs/1605.09757

[7] L. V. Gomez and J. Miura, “Ontology learning of new concepts combining textural knowledge and visual analysis for service robot applications,” in Proc. IEEE Int. Conf. Intell. Syst. Robot. (ISR), Mar. 2021, pp. 98–100.

[8] X. Jiang and A.-H. Tan, “CRCTOL: A semantic-based domain ontology learning system,” J. Amer. Soc. Inf. Sci. Technol., vol. 61, no. 1, pp. 150–168, 2010. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21231

[9] L. Karoui, M.-A. Aufaure, and N. Bennacer, “Contextual concept discovery algorithm,” in Proc. FLAIRS Conf., 2007, pp. 460–465.

[10] E. Drymonas, K. Zervanou, and E. G. M. Petrakis, “Unsupervised ontology acquisition from plain texts: The OntoGain system,” in Natural Language Processing and Information Systems, C. J. Hople, Y. Rezagui, E. Métais, A. Preece, and H. Li, Eds. Berlin, Germany: Springer, 2010, pp. 277–287.

[11] B. Frikh, A. Djaanfar, and B. Ouhbi, “A hybrid method for domain ontology construction from the web,” in Proc. KEOd, Jan. 2011, pp. 285–292.

[12] S. Chuprina, V. Alexandrov, and N. Alexandrov, “Using ontology engineering methods to improve computer science and data science skills,” Proc. Comput. Sci., vol. 80, pp. 1780–1790, Jan. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211064615324770

[13] A. Salatino, T. Thanhpalasingam, A. Mannocci, F. Osborne, and E. Motta, “The computer science ontology: A large-scale taxonomy of research areas,” in Proc. Int. Semantic Web Conf., Jan. 2018, pp. 187–205.

[14] R. C. Jackson, J. P. Balhoff, E. Douglass, N. L. Harris, C. J. Mungall, and J. A. Overton, “ROBOT: A tool for automating ontology workflows,” BMC Bioinf., vol. 20, no. 1, pp. 1–10, Dec. 2019.

[15] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. Goldberg, K. Eilbeck, A. Ireland, C. Mungall, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.-A. Sansone, R. Scheuermann, N. Shah, P. Wheitzel, and S. Lewis, “The OBO foundry: Coordinated evolution of ontologies to support biomedical data integration,” Nature Biotechnol., vol. 25, pp. 1251–1255, Nov. 2007.

[16] T. Nakamura, T. Araki, T. Nagai, and N. Iwahashi, “Grounding of word meanings in latent Dirichlet allocative based multimodal concepts,” Adv. Robot., vol. 25, no. 17, pp. 2189–2206, Jan. 2011.
[17] A. Ayub and A. R. Wagner, “Tell me what this is: Few-shot incremental object learning by a robot,” CoRR, vol. abs/2008.00819, 2020. [Online]. Available: https://arxiv.org/abs/2008.00819 and https://dblp.org/rec/journals/corr/abs-2008-00819.bib

[18] G. A. Miller, “WordNet: A lexical database for English,” Commun. ACM, vol. 38, no. 11, pp. 39–41, Nov. 1995, doi: 10.1145/219717.219748.

[19] H. Cunningham, V. Tablan, A. Roberts, and K. Bontcheva, “Getting more out of biomedical documents with GATE’s full lifecycle open source text analytics,” PLOS Comput. Biol., vol. 9, no. 2, pp. 1–16, Feb. 2013, doi: 10.1371/journal.pcbi.1002854.

[20] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” in Proc. Eur. Conf. Comput. Vis., 2016, pp. 630–645.

[21] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet large scale visual recognition challenge,” Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, 2015.

[22] L. V. Gómez and J. Miura, “Ontology-based knowledge management with verbal interaction for command interpretation and execution by home service robots,” Robot. Auton. Syst., vol. 140, Jun. 2021, Art. no. 103763. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921889021000488

LILIANA VILLAMAR GÓMEZ received the B.Eng. degree in computer systems engineering from Universidad del Valle de México, Mexico, in 2014, and the M.Eng. degree in information engineering from the Toyohashi University of Technology, Japan, in 2018, where she is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Graduate School of Engineering. Her research interests include service robots, human–robot interaction, artificial intelligence, and language processing.

JUN MIURA (Member, IEEE) received the B.Eng. degree in mechanical engineering, the M.Eng. and Dr. Eng. degrees in information engineering from the University of Tokyo, Tokyo, Japan, in 1984, 1986, and 1989, respectively. In 1989, he joined the Department of Computer-Controlled Mechanical Systems, Osaka University, Suita, Japan. Since April 2007, he has been a Professor with the Department of Computer science and Engineering, Toyohashi University of Technology, Toyohashi, Japan. From March 1994 to February 1995, he was a Visiting Scientist with the Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA. He published over 200 papers in international journals and conferences in the areas of intelligent robotics, mobile service robots, robot vision, and artificial intelligence. He received several awards, including the Best Paper Award from the Robotics Society of Japan, in 1997, the Best Paper Award Finalist at ICRA-1995, and the Best Service Robotics Paper Award Finalist at ICRA-2013.