University of Groningen

From Object Detection to Room Categorization in Robotics

Fernandez-Chaves, David; Ruiz-Sarmiento, Jose Raul; Petkov, Nicolai; Gonzalez-Jimenez, Javier

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

From Object Detection to Room Categorization in Robotics

David Fernandez-Chaves^{a,b}, Jose-Raul Ruiz-Sarmiento^a, Nicolai Petkov^b, Javier Gonzalez-Jimenez^a

^{a} Machine Perception and Intelligent Robotics group (MAPIR). Dept. of System Engineering and Automation.
Biomedical Research Institute of Malaga (IBIMA). University of Malaga. Spain
{davfercha,jotaraul,javiergonzalez}@uma.es

^{b} Johann Bernoulli Institute of Mathematics and Computing Science, University of Groningen, The Netherlands
n.petkov@rug.nl

ABSTRACT

This article deals with the problem of room categorization, *i.e.* the classification of a room as being a bathroom, kitchen, living-room, bedroom, etc., by an autonomous robot operating in home environments. For that, we propose a room categorization system based on a Bayesian probabilistic framework that combines object detections and its semantics. For detecting objects we resort to a state-of-the-art CNN, *Mask R-CNN*, while the meaning or semantics of those detections is provided by an ontology. Such an ontology encodes the relations between object and room categories, that is, in which room types the different object categories are typically found (toilets in bathrooms, microwaves in kitchens, etc.). The Bayesian framework is in charge of fusing both sources of information and providing a probability distribution over the set of categories the room can belong to. The proposed system has been evaluated in houses from the Robot@Home dataset, validating its effectiveness under real-world conditions.

CCS CONCEPTS

• **Computing methodologies** → **Cognitive robotics**; **Probabilistic reasoning**; *Ontology engineering*.

KEYWORDS

Room Categorization, Mobile Robots, Semantic Knowledge, Bayesian Inference, Ontologies, Uncertainty Propagation, Object Recognition

1 INTRODUCTION

Intelligent robots need to acquire and manage high-level information about their workspace in order to understand and successfully accomplish human commands like "go to the kitchen" or "bring me

Figure 1: Categorization of a room by a mobile robot according to the objects detected therein and its semantics.

the red book from the bedroom". Homes are usually split into functional areas, that we call rooms, where different human activities take place, like cooking, resting, having fun, etc. For a mobile robot, the ability to identify the category of a room: bedroom, kitchen, bathroom, living-room, etc., opens the door to a more comprehensive understanding of its workspace as well as to the possibility to carry out a wider variety of tasks [1, 19].

Different approaches can be found in the literature facing the *room categorization* problem. For example, there are works that employ classifiers based on geometric or appearance features of the room to categorize it [9, 10]. Other works exploit the fact that, given that rooms have a certain functionality, the objects they contain represent valuable information for their categorization [2, 8]. For instance, since kitchens are spaces for storing and preparing food, refrigerators or microwaves are usually placed there, while beds are typically in bedrooms and couches in living-rooms. This kind of *semantic knowledge* permits the utilization of detected objects as a hint to infer the category of the room [13, 14]. From the point of view of a mobile robot, which usually also needs to manage information about their surrounding objects, this approach results specially effective.

A clear drawback of the latter approach is that the performance of the room categorization system highly depends on the reliability of the object detection method. For example, if a night stand is misclassified as a microwave, the room could be categorized as a kitchen instead of as a bedroom. In recent years, *Convolutional Neural Networks* (CNNs) like Faster R-CNN [12], YOLOv3 [11] or

Mask R-CNN [5], have considerably increased the reliability of object recognition systems, but wrong detections may still appear producing incoherent room categorization results.

In this paper we contribute a room categorization system that combines in a novel way object detections and its semantics, pursuing a more coherent performance in the mobile robotics context. At the core of this system is a Bayesian probabilistic framework, which outputs a probability distribution over the set of categories a room can belong to. For that end, such a framework is fed with both: the objects detected in a sequence of images from a camera mounted on a robot, as well as the meaning or semantics of those detections. For recognizing objects we have resorted to Mask R-CNN, a state-of-the-art network that provides masks over the inspected images delimiting the pixels belonging to the detected objects, as well as scoring values measuring its confidence about such detections. In the proposed system multiple detections of the same object are fused, hence increasing their consistency. Regarding the meaning of these detections, it is retrieved from the *Semantic Knowledge* about categories of rooms and objects, as well as their relations, which are codified into an ontology [14, 20]. The consideration of the uncertainty about the object recognition results, their fusion when they belong to the same physical objects, as well as the exploitation of their semantics, allow the proposed system to achieve coherent categorization results. Fig. 1 illustrates an example of the outcome of our system while a robot is inspecting a room and some objects have been detected.

The proposed system is further described in Sec. 2. To assess the suitability of our approach when running on a mobile robot, we have carried out several experiments in Sec. 3 with the Robot@Home dataset [17]. This dataset gives us data from RGB-D cameras [22], a laser scanner and robot location along its trajectory in different houses. We conclude the paper in Sec. 4 discussing the work done and its possible extensions.

2 SYSTEM DESCRIPTION

This section describes the proposed system to categorize rooms by exploiting the objects detected therein and its semantics. Fig. 2 outlines its pipeline. Briefly, a CNN is used to recognize the objects in the room from RGB images, which returns the class, a pixel mask and a confidence score for each detected object. Next, the robot localization in the geometric (global) map of the environment is used to compose its pose and the pose of the object, expressed in the robot local reference frame. Since we are using calibrated RGB-D cameras [21, 22], this information is available in the depth image. Such composition permits to locate the object in such a geometric map, so it can be checked if the object was previously detected and if it is contained in the room being inspected. The relevant information about the object is added to an ontology (Sec. 2.2), which also contains previously stored human knowledge (HK) describing the different categories of rooms and objects that can be found in a house, together with their relations (in our case, in which room categories usually appear the considered object categories: microwaves in kitchens, coaches in living-rooms, etc.).

Through high-level queries, we retrieve the objects that have been previously detected inside a certain room and their scores, as well as the probability for each class object appearing in each room
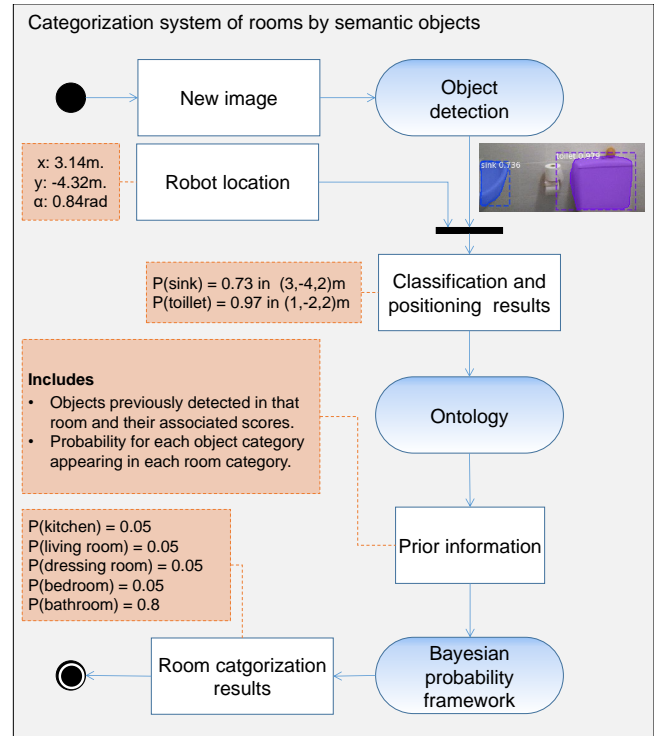
Figure 2: Activity graph (employing unified modeling language (UML)) showing the workflow of the system used to categorization system of rooms by semantic objects.

category. Finally, a Bayesian probabilistic framework is in charge of combining such information to compute a probability distribution over the categories the inspected room can belong to (Sec. 2.3).

The following sections introduce the three main components of the proposed system: object detection, semantic information and room category inference.

2.1 Object detection

One of the fundamental components of our system is the object detector, whose function is to analyze RGB images from a camera mounted on a robot to obtain information about the objects in its surroundings. For that purpose, we rely on a popular CNN, namely Mask R-CNN [5], which yields: i) the class of the detected objects, ii) masks of the pixels that belong to each class, and iii) confidence scores. Fig. 2 shows at the output of the *object detector* a fragment of an image with overlapping pixel masks indicating that a sink and a toilet have been detected.

From these pixel masks, the position of a detected object in the robot local frame can be retrieved in different ways. In our case, since we make use of (extrinsically and intrinsically) calibrated RGB-D cameras, such position can be extracted from the depth image in the form of point clouds. Then, through the composition of the robot pose in the (global) geometric map of the house and the object position, the position of the object in such a map can be obtained. This approach was also followed in [15]. The global location of the detected object is useful for placing it within a certain
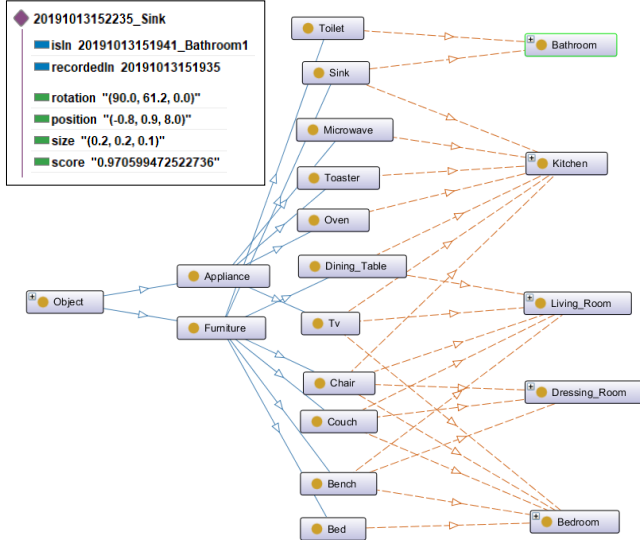
Figure 3: Extract from the ontology used in this work, where the blue arrows represent the *isA* predicate and the orange ones stand for the *isIn* predicate. Top left, example of the description in the ontology of an instance of the concept *Sink*, where its relations with other elements are marked in blue, and its properties in green.

room as well as for checking future detections of the same object. In this work, we consider that two observations refer to the same object when they have the same category and the distance between the position of both is less than a given threshold. This allows the system to integrate knowledge about object detections over time, reducing the uncertainty when objects are observed multiple times.

2.2 Semantic Knowledge

The usage of semantic knowledge encoded in ontologies exhibits significant advantages for robots in a variety of tasks [3, 14, 18]. An ontology can be defined as a formal representation of the concepts related to a domain and their interrelations [20]. In this paper we work with the *house* domain, the concepts are the categories of objects (*e.g. chair*, *bed*, *book*, etc.) and rooms (*e.g. kitchen*, *bedroom*. *bathroom*, etc.), and their relations are codified through predicates like *isA* or *isIn*. With these resources, we can create instances of objects that belong to a category (*e.g. isA(chair_2020,chair)*) and relate them to other categories (*e.g. isIn(chair_2020,kitchen)*) or properties (*e.g. score(chair_2020, 0.88)*).

Semantic knowledge can be acquired in different ways, being *human elicitation* [16] the choice in this work. Fig. 3 shows a fragment of the resultant ontology codifying such knowledge, where the relations between categories of objects and rooms are represented as orange arrows. During the robot operation in the house, every time a new object is detected, a new instance is created in the ontology that is described through a number of predicates. Fig. 3-top-left illustrates an example of an object instance, including: i) the category of the object as reported by the CNN, ii) the properties

of the object (marked with green squares), and iii) the room where it has been seen according to its position (blue squares).

To implement such an ontology we have resorted to RDFSharp[1], which allows us to use the Web Ontology Language (OWL) to encode the described predicates. In addition, this software package supports the execution of high-level queries through SPARQL, giving us the facility to retrieve the information required to feed the next component of the categorization system (*e.g. Give me all the objects in this room (bathroom_1024)*).

2.3 Bayesian Probabilistic Framework

To carry out the pursued categorization of rooms we have designed a Bayesian framework [7] able to combine and exploit both sources of information: the objects previously detected in the room (including confidence scores about them), and the meaning or semantics of these detections. In short, this framework formally performs reasoning like the following one "a microwave and an oven have been detected with a high level of confidence, so this room must be a kitchen". The output of this component is a probability distribution over the set of possible classes the room can belong to (bathroom, bedroom, kitchen, livingroom, etc.). Let us introduce the following definitions in order to properly state the problem:

- Let n be the number of objects detected in the room.
- n_{oc} represents the number of considered object categories.
- n_{rc} models the number of considered room categories.
- Define $\mathbf{z}_i = [z_{i,1}, \ldots, z_{i_m}], i = 1 : n$ as a a vector containing the m (visual) observations of the object i.
- $O_i = \{O_i^j, j = 1 : n_{oc}\}$ is a random variable modeling the category of object i, taking values on the set of possible object categories OC.
- R is a random variable classifying the room by taking values on the set of possible room categories RC.

Initially, when no information about the objects in the room is available, the probability for it belonging to a certain category $P(R = RC_i)$ (simplified as $P(R)$ for the sake of clarity) is defined as a uniform distribution, that is:

$$P(R) = 1/n_{rc} \qquad (1)$$

For a better understanding of the proposed formulation, let us start by describing a simple case where a unique object is detected in the room through the set of observations $\mathbf{z}$. Then, the probability in Eq. (1) is modified according to the output of the object detection and semantic knowledge components as follows:

$$P(R|\mathbf{z}) = \sum_{i=1}^{N_{oc}} P(R|\mathbf{z}, O^i)P(O^i|\mathbf{z}) \qquad (2)$$

that is, such a probability is obtained by marginalizing over the possible object categories that the detected object could belong to. In such a marginalization two probabilities appear. The first one, $P(R|\mathbf{z}, O^i)$, represents the probability that the room belongs to a certain category conditioned on the object observations $\mathbf{z}$ and the object category O^i. Since we can safely assume that $R \perp\!\!\!\perp \mathbf{z} \mid O^i$, *i.e.* that the category of the room is independent of the object observations given the object category, it can be simplified to $P(R|O^i)$.

---

[1] https://www.w3.org/2001/sw/wiki/RDFSharp

This probability is defined by the information codified in the ontology, setting which object categories may appear in which room categories. In this way, this probability distribution is defined as:

$$P(R|O^i) = \begin{cases} 0.9/count(O^i, RC) & \text{if } \exists \; isIn(O^i, R) \\ 0.1/(n_{rc} - count(O^i, RC)) & \text{if } \nexists \; isIn(O^i, R) \end{cases} \quad (3)$$

being $count(O^i, RC)$ a function returning the number of room categories in which the object category O^i may appear, and $isIn(O^i, R)$ the predicate introduced in Sec. 2.2. Tab. 1 shows an example of the computation of this distribution for some object and room categories.

The second probability distribution, $P(O^i|\mathbf{z})$, models the probability that an object belongs to a category given its observations. This probability is obtained from the ontology according to the detection results. Concretely, Mask R-CNN is configured to return detections with a confidence score in the range [0.7,1], and the information relative to their pose, size and score are sent to the ontology. In this way, when an object is detected, the information already present in the ontology is used to check if it had been previously observed. If so, we fuse those detections in order to work with more robust and coherent information. This is done in a simple but effective fashion: by averaging the yielded objects poses, sizes, and scores. Such averaging is weighted by the relevance of each detection, measured by their individual scores, so more confident detections contribute to a larger extent in this data fusion process. Finally, in order to have a probability distribution over the set of possible object categories OC, we consider the object category score, assign a score of 0.1 for the remaining categories, and normalize these values by dividing by the total sum of scores.

Once we have described how to compute the probability distribution over room categories given a detected object, let us extend this formulation handle multiple objects. Concretely, this probability is retrieved by recursively marginalizing over all the detected objects, that is:

$$P(R|\mathbf{z}_1, \ldots, \mathbf{z}_n) = \sum_{i_1=1}^{N_{oc}} P(R|O_1^{i_1}, \mathbf{z}_1, \ldots, \mathbf{z}_n) P(O_1^{i_1}|\mathbf{z}_1)$$

$$P(R|O_1^{i_1}, \mathbf{z}_2, \ldots, \mathbf{z}_n) = \sum_{i_2=1}^{N_{oc}} P(R|O_1^{i_1}, O_2^{i_2}, \mathbf{z}_2, \ldots, \mathbf{z}_n) P(O_2^{i_2}|\mathbf{z}_2)$$

$$\cdots$$

$$P(R|O_1^{i_1}, \ldots O_n^{n_1}) = \sum_{i_n=1}^{N_{oc}} P(R|O_1^{i_1}, \ldots, O_n^{i_n}) P(O_n^{i_1}|\mathbf{z}_n)$$

$$(4)$$

All the probabilities in this definition have been previously introduced, except $P(R|O_1^{i_1}, \ldots, O_n^{i_n})$. By applying the Bayes rule to it, and assuming that $RO_i \perp\!\!\!\perp O_j \mid R$, it can be expressed as:

$$P(R|O_1^{i_1}, \ldots, O_n^{i_n}) = \prod_{m=1}^{n} P(R|O_m^{i_m}) \quad (5)$$

This recursive model can be expensive to calculate when there are many objects within the room. In order to prevent this, the probabilistic framework is only fed with the N objects having more detections. This approach also performs as a filter that ignores objects with spurious detections, which are probably wrong.

3 SYSTEM EVALUATION

To assess the performance of the proposed room categorization system we have carried out an experiment where a mobile robot has to operate in real houses. For that, we have resorted to the

Table 1: Conditional room category probabilities for each object category: $P(R|O^i)$. As can be seen, the probabilities used depend on the number of categories in which each object can be detected following Eq. (3)

| Object Class	Kitchen	Living room	Dressing room	Bedroom	Bathroom
Microwave	0.9	0.025	0.025	0.025	0.025
Oven	0.9	0.025	0.025	0.025	0.025
Toaster	0.9	0.025	0.025	0.025	0.025
Tv	0.3	0.3	0.05	0.3	0.05
Bed	0.025	0.025	0.025	0.9	0.025
Bench	0.05	0.3	0.3	0.3	0.05
Chair	0.225	0.225	0.225	0.225	0.1
Couch	0.05	0.3	0.3	0.3	0.05
Dining Table	0.45	0.45	0.03	0.03	0.03
Sink	0.45	0.03	0.03	0.03	0.45
Toilet	0.025	0.025	0.025	0.025	0.9

Figure 4: On the left, the Giraff robot used in Robot@Home, with a ring of 4 RGB-D cameras mounted. To the right, an example of images of intensity (above) and depth (below) of the dataset captured by the cameras.

Robot@Home dataset, described in Sec. 3.1. The setup of the experiment is presented in Sec. 3.2 and, finally, we comment on the obtained results (Sec. 3.3).

3.1 Dataset: Robot@Home

Robot@Home [17] is a repository of data gathered in real houses by a Giraff robot [4] (see Fig. 4-left). This mobile robot was equipped with a ring of four vertically positioned RGB-D cameras and a 2D laser scanner mounted on the base. The dataset provides us with different raids on real houses, also including geometric maps of those houses and the localization of the robot during the raids. Additionally, it also contains other processed data such as 3D reconstructions, used in this work for the visualization of the results, and segmented objects and rooms annotated with their ground truth categories. It is publicly available at: http://mapir.isa.uma.es/work/robot-at-home-dataset.

Fig. 4-right shows an example of four RGB-D observations simultaneously collected from the cameras onboard the robot. In order to show the generality and applicability of our proposal for room categorization through object detection, we have only considered images coming from the RGB-D camera looking forward, as it supposes a more common sensory configuration in mobile robots.
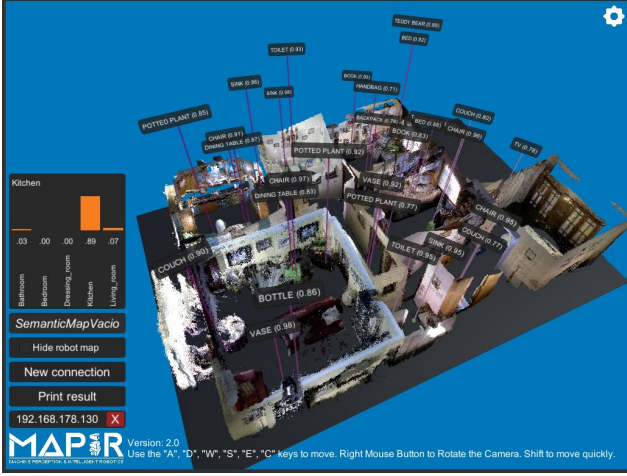
Figure 5: Virtual map of the house *Anto* with annotations of the detected objects. On the left, the system interface and a bar diagram representing the probability distribution over the different categories the room in which the robot is located can belong to.

3.2 Experimental Setup

In the experiment carried out the robot is tasked to categorize the rooms of four houses within Robot@Home, concretely: *alma*, *anto*, *pare* and *rx2*, considering five different categories: *Bathroom*, *Bedroom*, *Dressing Room*, *Kitchen* and *Living Room*.

For the object detection component an instance of Mask R-CNN pre-trained with the COCO [6] dataset is used, which includes object categories typically appearing in houses such as chair, couch, potted plant, bed, dining-table, toilet, tv monitor, etc. From these categories, in this experiment we take into account those that can be grouped into the *Furniture* and *Appliance* super-categories. The reason for this is that they are static and highly related to certain room categories, unlike more general objects like backpacks or books.

Regarding the semantic knowledge component (recall Sec. 2.2), as commented, Fig. 3 shows the relationships between the categories of objects and rooms used in this experiment, setting where they could appear. These relations are rendered into probabilities using Eq. (3), as illustrated in Tab. 1. Some object categories are exclusive to one category of room, which means that finding them in another category is unlikely (*e.g.* microwaves are usually in the kitchen). On the other hand, other objects are more general, like chairs, so they are less discriminant where categorizing rooms. Concerning the number of detected objects to be considered by the Bayesian framework, we have empirically checked that with $N = 5$ the system achieves a good trade-off between execution time and categorization results.

3.3 Experimental Results

Fig. 5 shows an example of the interface used in the integration of our system. We employ point clouds and virtual annotations to represent the knowledge of the robot as it explores the environment. This allows us to debug the performance of the system in a visual

Table 2: Probability distributions obtained for each room in *Anto* (rows), a house from the *Robot@Home* dataset. Columns represent the possible room categories. Correct categorizations are marked in green, in yellow those that are not decisive, and in red wrong ones.

Rooms	Kitchen	Living room	Dressing room	Bedroom	Bathroom
Kitchen-1	0.98	0.002	0.002	0.002	0.01
Living_room-1	0.025	0.025	0.025	0.91	0.025
Dressing_room-1	0.05	0.3	0.3	0.3	0.22
Bedroom-1	0.004	0.004	0.004	0.96	0.004
Bedroom-2	0.02	0.02	0.02	0.91	0.01
Bedroom-3	0.07	0.42	0.07	0.42	0.005
Bathroom-1	0.05	0.05	0.05	0.05	0.92
Bathroom-2	0.026	0.011	0.011	0.011	0.93

Table 3: Results obtained from 5 executions in each house. The last row shows the global average.

	#Rooms × executions	Success	Inconclusive	Wrong
Alma	5 × 5	0.60	0.20	0.20
Anto	8 × 5	0.60	0.35	0.03
Pare	8 × 5	0.58	0.30	0.13
Rx2	4 × 5	0.55	0.31	0.20
	Global Average	0.59	0.29	0.12

and friendly way, knowing at any given moment where and what detections the robot has made. In addition, the results obtained from the probabilistic framework are represented on the left of the interface in the form of a bar graph. Such results are updated each time the object detection component raises a detection.

For each execution of the experiment, the proposed categorization system retrieves, for every room, a probability distribution over the considered categories according to the objects detected therein. Tab. 2 shows an example of the results of one execution with *Anto* house. The first column shows the ground truth category of each room, while the other columns report the probabilities computed by the system for the considered categories. In this table the successful categorizations are highlighted in green, the wrong one in red, and the two rooms with inconclusive results in yellow. It is worth mentioning that the wrong categorization was due to the misclassification of a couch as a bed.

Tab. 3 provides the accuracy achieved by the categorization system, obtained from running it five times with the data collected from each house. This is done because the results between executions in the same house may differ depending on the images processed by the object detection component, since they are provided at a higher frequency than the one achieved by such component, which only processes the most recent one. The last row shows the average of success, inconclusive and wrong results for the 20 runs. We can see how, in general, the system reaches a high categorization accuracy, exhibiting only a 12% of wrong classifications. The interested reader can see the categorization system in action in the following video: https://youtu.be/0suNQ1v6uVU.

The confusion matrix obtained from the 20 executions of the system is depicted in Fig. 6. In such a matrix, rows index ground truth categories, while columns index the categories returned by the categorization system. From there we can verify that, regarding categorization accuracy, the three top categories are *Bathroom*

	Kitchen	Living room	Dressing room	Bedroom	Bathroom
Kitchen	0.77	0.08	0.08	0.08	0.00
Living room	0.21	0.33	0.10	0.37	0.00
Dressing room	0.06	0.31	0.31	0.31	0.00
Bedroom	0.29	0.05	0.00	0.65	0.00
Bathroom	0.03	0.00	0.00	0.00	0.97

Figure 6: Confusion matrix from the categorization of rooms in the conducted experiments.

(97%), *Kitchen* (77%) and *Bedroom* (65%), while there are some categories difficult to distinguish: *Living room* and *Dressing rooms*. This inconclusive categorizations are mainly due to the absence of detections of objects belonging to categories that typically appear in those room types. Results are more precise when there are detected objects that can be only found in a reduced set of room categories, *e.g.* when a toilet is detected multiple times, the probability of the room being a bathroom notably increases. Instead, detections of objects common to multiple categories, as is the case of chairs (may appear in living rooms, kitchens, dressing rooms or bedrooms), provide poor information towards room categorization.

After analyzing the results, we observed that misclassifications are mainly due to incorrect object detections that, combined with a reduced number of objects detected in the room, cause the system to perform incorrectly. A way to mitigate this issue could be the utilization of an object classifier able to detect a wider range of object categories, hence providing more detections towards a more robust performance. Besides, the experiment was carried out passively, i.e. while the robot was moving according to the path it followed during the dataset collection. We argue that the results could improve if the robot was able to carefully inspect each room in order to get further information (additional object detections), perhaps also including an active perception module.

4 CONCLUSIONS

This paper has presented a room categorization system able to combine and exploit object detections and its semantics. The proposed approach relies on the use of a state-of-the-art CNN to recognize objects, an ontology to encode the relationships between objects and room categories (semantic knowledge), and a Bayesian probabilistic framework to fuse all the information and provide a probability distribution over the set of categories the room can belong to. These components permit us to leverage the fact that objects are placed in rooms according to their functionality, as well as to manage the uncertainty inherent to the object detection and room categorization processes. The outcome of the system can be further exploited by mobile robots to efficiently perform high-level tasks.

The suitability of our proposal has been evaluated in four real houses from the Robot@Home dataset. The reported results support our claim that objects' detections and its meaning are valuable resources towards room categorization, only producing a 12% of wrong classifications.

As future work, we plan to design an algorithm to evaluate when the robot has gather enough information for the categorization to be conclusive. For this, rather than the detection being a passive process, the robot should carefully inspect every room looking for objects.

REFERENCES

[1] D. Chaves, J. R. Ruiz-Sarmiento, N. Petkov, and J. Gonzalez-Jimenez. 2019. Integration of CNN into a Robotic Architecture to Build Semantic Maps of Indoor Environments. In *Advances in Computational Intelligence*. 313–324.

[2] Pablo Espinace, Thomas Kollar, Alvaro Soto, and Nicholas Roy. 2010. Indoor scene recognition through object detection. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 1406–1413.

[3] Cipriano Galindo, Juan-Antonio Fernández-Madrigal, and Javier González-Jiménez. 2008. Multihierarchical Interactive Task Planning. Application to Mobile Robotics. *IEEE Transactions on Systems, Man, and Cybernetics, part B* 38, 3 (2008), 785–798.

[4] Javier González-Jiménez, Cipriano Galindo, and JR Ruiz-Sarmiento. 2012. Technical improvements of the Giraff telepresence robot based on users' evaluation. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 827–832.

[5] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask R-CNN. *arXiv e-prints*, Article arXiv:1703.06870 (Mar 2017). arXiv:cs.CV/1703.06870

[6] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. In *European Conference on Computer Vision*. 740–755.

[7] Javier Monroy, J. R. Ruiz-Sarmiento, Francisco-Angel Moreno, Francisco Melendez-Fernandez, Cipriano Galindo, and Javier Gonzalez-Jimenez. 2018. A Semantic-Based Gas Source Localization with a Mobile Robot Combining Vision and Chemical Sensing. *Sensors* 18, 12 (2018).

[8] Andrzej Pronobis and Patric Jensfelt. 2011. Hierarchical Multi-Modal Place Categorization.. In *European Conference on Mobile Robots (ECMR)*. 159–164.

[9] Andrzej Pronobis, O Martínez Mozos, Barbara Caputo, and Patric Jensfelt. 2009. Multi-modal semantic place classification. *The International Journal of Robotics Research* (2009).

[10] Ariadna Quattoni and Antonio Torralba. 2009. Recognizing indoor scenes. In *Computer Vision and Pattern Recognition, 2009. IEEE Conference on*. 413–420.

[11] Joseph Redmon and Ali Farhadi. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).

[12] S. Ren, K. He, R. Girshick, and J. Sun. 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 6 (June 2017), 1137–1149.

[13] J. G. Rogers and H. I. Christensen. 2012. A conditional random field model for place and object classification. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. 1766–1772.

[14] J. R. Ruiz-Sarmiento, Cipriano Galindo, and Javier González-Jiménez. 2015. Joint Categorization of Objects and Rooms for Mobile Robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

[15] J. R. Ruiz-Sarmiento, Cipriano Galindo, and Javier González-Jiménez. 2015. OLT: A Toolkit for Object Labeling Applied to Robotic RGB-D Datasets. In *European Conference on Mobile Robots*.

[16] J. R. Ruiz-Sarmiento, Cipriano Galindo, and Javier González-Jiménez. 2017. Building Multiversal Semantic Maps for Mobile Robot Operation. *Knowledge-Based Systems* 119 (2017), 257–272.

[17] J. R. Ruiz-Sarmiento, Cipriano Galindo, and Javier González-Jiménez. 2017. Robot@Home, a Robotic Dataset for Semantic Mapping of Home Environments. *The International Journal of Robotics Research* 36, 2 (2017), 131–141.

[18] J. R. Ruiz-Sarmiento, Cipriano Galindo, Javier Monroy, Francisco-Angel Moreno, and Javier Gonzalez-Jimenez. 2019. Ontology-based conditional random fields for object recognition. *International Journal of Knowledge-Based Systems* 168 (mar 2019), 100–108.

[19] P. UrÅaiÄŊ, R. Mandeljc, A. Leonardis, and M. Kristan. 2016. Part-based room categorization for household service robots. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*. 2287–2294.

[20] Mike Uschold and Michael Gruninger. 1996. Ontologies: principles, methods and applications. *The Knowledge Engineering Review* 11 (1996), 93–136. Issue 02.

[21] David ZuÃśiga-NoÃŋl, J. R. Ruiz-Sarmiento, Ruben Gomez-Ojeda, and Javier Gonzalez-Jimenez. 2019. Automatic Multi-Sensor Extrinsic Calibration For Mobile Robots. *IEEE Robotics and Automation Letters* 4, 3 (jul 2019), 2862–2869.

[22] David ZuÃśiga-NoÃŋl, J. R. Ruiz-Sarmiento, and Javier Gonzalez-Jimenez. 2019. *Intrinsic Calibration of Depth Cameras for Mobile Robots using a Radial Laser Scanner*. CAIP. Lecture Notes in Computer Science, Vol. 11678. 659–671.