Mobile Robot Object Recognition through the Synergy of Probabilistic Graphical Models and Semantic Knowledge

J.R. Ruiz-Sarmiento and **C. Galindo** and **J. Gonzalez-Jimenez**[1]

Abstract. Mobile robots intended to perform high-level tasks have to recognize objects in their workspace. In order to increase the success of the recognition process, recent works have studied the use of contextual information. Probabilistic Graphical Models (PGMs) and Semantic Knowledge (SK) are two well-known approaches for dealing with contextual information, although they exhibit some drawbacks: the PGMs complexity exponentially increases with the number of objects in the scene, while SK are unable to handle uncertainty. In this work we combine both approaches to address the object recognition problem. We propose the exploitation of SK to reduce the complexity of the probabilistic inference, while we rely on PGMs to enhance SK with a mechanism to manage uncertainty. The suitability of our method is validated through a set of experiments, in which a mobile robot endowed with a Kinect-like sensor captured 3D data from 25 real environments, achieving a promising result of $\sim$94% of success.

1 INTRODUCTION

A mobile robot intended to perform high-level tasks has to be aware of its surrounding. Recent works have studied the use of the objects' contexts in order to exploit their spatial relations [1, 2, 3, 4]. This kind of information is crucial to disambiguate recognition results where appearance and geometric features are not discriminative enough [5]. For example, an object recognition system could not properly identify a cylindrical object as a lamp or a trash bin, but if it is found on the floor and near a wall, the trash bin option should stand out as the most likely one.

Probabilistic Graphical Models (PGMs) [6] exploit contextual relations, and have been recently used for object recognition with acceptable success [1, 2, 3, 7]. However, the complexity of the probabilistic inference process, which consists in finding the most probable class assignation to a set of objects, becomes intractable when the number of objects and classes augments. For example, a scene with 10 objects, which can belong to 9 different classes, e.g. table, chair, computer screen, etc., entails 9^{10} possible assignations, and although approximated inference methods can be used to reduce the search space [8, 9, 10], the overall performance is compromised.

An alternative to PGMs is the utilization of semantic knowledge, which can be naturally codified in the form of object classes (e.g. `Table`, `Chair`, etc.), relations between classes (e.g. `Table isNear Chair`) and instances of them (e.g. `table-1`). One of the advantages of these definitions is that they are common-sense and human-readable, facilitating in this way the information exchange between robots and humans. However, although this methodology has been used to recognize objects through logical reasoners [11, 12], they are unable to handle uncertainty, and it is difficult to leverage all the potential of the contextual relations.

In this paper we present an approach that combines semantic knowledge and probabilistic graphical models to address the object recognition problem in mobile robot applications. Concretely, we propose the use of ontologies [13] to codify SK by mean of expert elicitation, and its exploitation for object recognition through Conditional Random Fields (CRF) [6], which are trained using synthetic samples as described in [14]. In this combination, CRFs provide the ontology with a probabilistic reasoning mechanism that handles uncertainty and contextual relations, while the semantics contributes with:

- *A significant reduction in the complexity of the probabilistic inference*. The ontology is used to generate hypotheses about the most probable class assignations for a given object according to its features, reducing thus the number of assignations. For example, the hypotheses yielded by the semantics for a vertical planar surface over the floor could be a `Wall`, a `Computer_screen` or a `Chair_back_rest`, but not a `Chair_seat`.
- *Prior information about the occurrence of objects*. Ontologies are a natural source of prior information when encoding the frequency of occurrence of the objects in a scene. For example, an ontology can encode common sense knowledge stating that a typical office should contain a computer, less probably a coach, but never a bathtub. We propose the use of this knowledge as prior information and present a modification of the usual CRF formulation to cope with it.

In this work, we consider planar patches extracted from RGB-D point clouds of the scene as the constituent parts of the objects to be recognized. The suitability of our method has been validated through a number of tests with real data gathered by a mobile robot from 25 office scenarios. An additional advantage of our approach is that the recognized objects are anchored to classes defined into the ontology, and this is useful for robotic high-level processes like reasoning or task planing [12, 15, 16].

Next we revise Conditional Random Fields as a tool for object recognition. Section 3 describes how semantic knowledge is applied to assist and improve the recognition process of a mobile robot. In section 4, the results of the method evaluation are shown. Finally, section 5 presents some conclusions of this work.

[1] Authors are with the System Engineering and Automation Dept., University of Málaga, Spain. http://mapir.isa.uma.es/. Corresponding author: `jotaraul@uma.es`

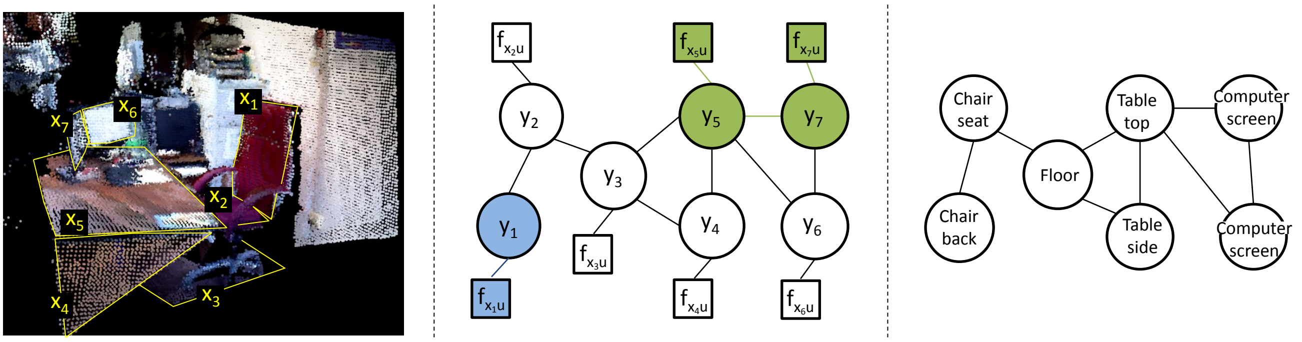

Figure 1. Left, planar patches x extracted from a scene, delimited by yellow lines. Middle, CRF built for that scene, where each y_i is associated with its respective x_i, and conditioned by its extracted features $\boldsymbol{f_{x_i u}}$. Blue shapes represent an example of the scope of an unary factor, while green ones the scope for a pairwise factor. Right, recognition result obtained through probabilistic reasoning over the CRF.

2 BACKGROUND ON SCENE OBJECT RECOGNITION THROUGH CONDITIONAL RANDOM FIELDS

From a probabilistic stance, the object recognition process can be formulated as follows. Let's have a scene with $\boldsymbol{x} = [x_1, .., x_n]$ observed objects (see figure 1-left), each one characterized by a vector of features $\boldsymbol{f}_{x_i u} = [f_{x_i u_1}, .., f_{x_i u_m}]$ (e.g. height, area, etc.), and $L = \{l_1, .., l_k\}$ the set of considered classes. Let $\boldsymbol{y} = [y_1, y_2, .., y_n] \mid y_i : \boldsymbol{L}^k \to \{0, 1\}^k$ be a vector of discrete random variables corresponding to the class assignment to **x**. Thus, the recognition process consists in maximizing the joint probability distribution $P(\boldsymbol{y}, \mathbf{x})$, i.e., to find the most probable classes assignation to $\boldsymbol{y}$, also maximizing a number of probability distributions over the features extracted from $\boldsymbol{x}$. Such a joint distribution has a high dimensionality, so its exhaustive definition is prohibitive. Probabilistic Graphical Models (PGMs) permits to break down such a definition into smaller pieces exploiting the concept of independence [6]. To simplify more the problem, we employ a particular type of PGM called Conditional Random Field (CRF) [6], which factorizes the distribution $P(\boldsymbol{y}|\boldsymbol{x})$, instead of encoding the probability distribution $P(\boldsymbol{y}, \boldsymbol{x})$. This avoids the definition of the probability distributions over the object features extracted from $\boldsymbol{x}$, which usually exhibits complex dependencies.

In general, a CRF is represented through a graph $H = \{V, E\}$, built upon two elements: a set of nodes V, and a set of edges E. Nodes V represent random variables, and edges E link nodes that keep some kind of relation, i.e., they are dependent. Typically, in visual object recognition, the nodes correspond to the random variables $\boldsymbol{y}$, and two nodes y_i and y_j are connected if their associated objects x_i and x_j are close in the scene (see figure 1-middle). The rationale of this is that the recognition of an object condition the recognition of nearby objects, but not those far away.

According to the Hammersley-Clifford theorem [6], the distribution $P(\boldsymbol{y}|\boldsymbol{x})$ can be factorized over H as a product of factors, being a factor a function that represents a probability distribution over a part of H. In this work we use two kinds of factors: local and pairwise. *Local factors* refer to nodes, and express how probable is that an observed object x_i belongs to a certain class from L according to its extracted features. On the other hand, *pairwise factors* are associated to pairs of nodes, and codify the compatibility of the classes assigned to a given pair.

Concretely, we define an unary factor, denoted by $U(\cdot)$, as a linear classification model:

$$U(y_i, x_i, \boldsymbol{\omega}) = \sum_{l \in L} \delta(y_i = l) \boldsymbol{\omega}_l f(x_i) \tag{1}$$

where $f(x_i)$ is the function in charge of computing the features $\boldsymbol{f}_{x_i u}$ for the object x_i, $\boldsymbol{\omega}_l = [\omega_{1,l}, .., \omega_{f_m,l}]$ is a vector of weights for each class $l \in L$ obtained during the training phase, and $\delta(y_i = l)$ is the Kronecker delta function, which takes value 1 when $y_i = l$ and 0 otherwise. Table 1-left shows the unary features used in this work.

On the other hand, a pairwise factor $I(\cdot)$ is defined as:

$$\begin{aligned} I(y_i, y_j, x_i, x_j, \boldsymbol{\theta}) \\ = \sum_{l_1 \in L} \sum_{l_2 \in L} \delta(y_i = l_1) \delta(y_j = l_2) \boldsymbol{\theta}_{l_1, l_2} g(x_i, x_j) \end{aligned} \tag{2}$$

where the function $g(x_i, x_j)$ computes a set pairwise features $\boldsymbol{f}_{x_i x_j p} = [f_{x_i x_j p_1}, .., f_{x_i x_j p_q}]$ for the relation between objects x_i and x_j, and $\boldsymbol{\theta}_{l_1, l_2} = [\theta_{1, l_1, l_2}, ...\theta_{q, l_1, l_2}]$ is a vector of weights for each pair of classes in L. In this work the CRF training, i.e., the estimation of the vectors of weights $\boldsymbol{\omega}$ and $\boldsymbol{\theta}$, is performed through the optimization of the pseudo-likelihood function [6].

For convenience, the factorization of $P(\boldsymbol{y}|\boldsymbol{x})$ over the graph H is expressed by means of log-linear models as:

$$P(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\omega}, \boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{x}, \boldsymbol{\omega}, \boldsymbol{\theta})} e^{-\epsilon(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\omega}, \boldsymbol{\theta})} \tag{3}$$

where $Z(\cdot)$ is the normalizing partition function so $\sum_{\xi(\boldsymbol{y})} p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\omega}, \boldsymbol{\theta}) = 1$, being $\xi(\boldsymbol{y})$ an assignation to the variables in $\boldsymbol{y}$, and $\epsilon(\cdot)$ the so-called energy function defined as:

$$\epsilon(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\omega}, \boldsymbol{\theta}) = \sum_{i \in V} U(y_i, x_i, \boldsymbol{\omega}) + \sum_{(i,j) \in E} I(y_i, y_j, x_i, x_j, \boldsymbol{\theta}) \tag{4}$$

Given an observation of the scene, the CRF graph $H = \{V, E\}$ is built according to the observed objects $\boldsymbol{x}$ and their proximity (objects at a distance below a given threshold are linked together), which set the conditional dependencies between the random variables in $\boldsymbol{y}$. Thus, the object recognition problem is that of finding the assignation to $\boldsymbol{y}$ that maximizes the posterior, that is:

$$\begin{aligned} \hat{y} &= arg \max_{y} P(\mathbf{y}|\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\theta}) \\ &= arg \max_{y} \frac{1}{Z(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\theta})} e^{-\epsilon(\mathbf{y}, \mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\theta})} \end{aligned}$$

Table 1. Unary and pairwise features used to characterize a planar patch and its relations.

id	**Unary features**
$f_{x_i u_1}$	Centroid height from the floor.
$f_{x_i u_2}$	Orientation w.r.t. the horizontal.
$f_{x_i u_3}$	Area of its bounding box.
$f_{x_i u_4}$	Elongation.
id	**Pairwise features**
$f_{x_i x_j p_1}$	Perpendicularity.
$f_{x_i x_j p_2}$	on/under relation.
$f_{x_i x_j p_3}$	Vertical distance of centroids.
$f_{x_i x_j p_4}$	Ratio between areas.
$f_{x_i x_j p_5}$	Ratio between elongations.

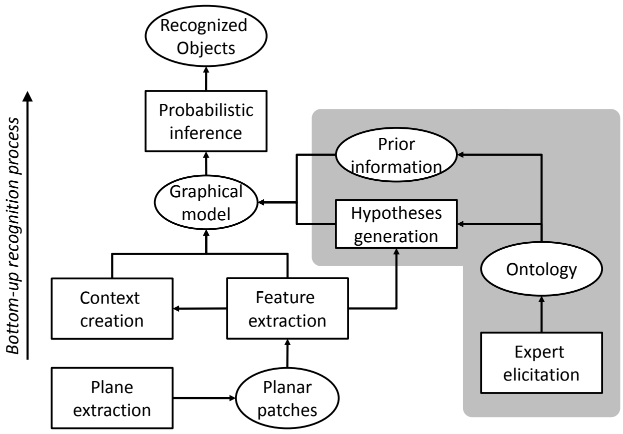

Figure 2. Overview of the proposed method for object recognition. Boxes are processes, while ovals represent generated/consumed data by the processes. The shadowed area encapsulates the components that directly make use of semantic knowledge.

Given that the partition function does not depend on the assignments to $\boldsymbol{y}$, such expression can be simplified by:

$$\hat{y} = arg \max_{y} e^{-\epsilon(\mathbf{y}, \mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\theta})} \tag{5}$$

This equation is known as the Maximum a Posteriori (MAP) or Most Probable Explanation (MPE) problem. Despite of the avoided computation of the partition function, its calculus is still unfeasible given that the number of possible assignations grows exponentially with the number of nodes in V, i.e., the number of objects to be recognized. Approximated inference methods can mitigate this problem, although their performance can be also compromised if the number of objects and classes is high. Particularly, during the evaluation of the proposed method we have relied on the Iterated Conditional Modes (ICM) [8] algorithm, which maximizes local conditional probabilities instead of the whole $P(\boldsymbol{y}|\boldsymbol{x})$. Figure 1-right shows, for the scene objects in figure 1-left, the most probable classes assignation computed by such a method.

3 EXPLOITATION OF SEMANTIC KNOWLEDGE

The recognition framework proposed in this work follows a bottom-up methodology (see figure 2). During the robot operation, 3D observations gathered from a Kinect-like sensor are registered [17], and planar patches are extracted and characterized through the set of features showed in table 1. This information is exploited by the ontology to hypothesize the most probable class assignments for each patch by means of logical inference[2]. These hypotheses dramatically reduce the number of potential classes to be considered by the CRF. Additionally, a modification to the usual CRF formulation has been carried out in order to also take advantage of prior information about the frequency of occurrence of the different object classes. In summary, the result of the recognition process is provided by probabilistic reasoning over a CRF, managing (i) a number of characterized planar patches from the scene, (ii) hypothesis about the most probable classes of each patch, and (iii) prior information about the occurrence of classes.

Next, section 3.1 describes the codification of semantic information through ontologies. Section 3.2 presents the use of an ontology to provide hypothesis, and section 3.3 introduces the utilization of prior information in a CRF.

3.1 Ontology Codification

An ontology is commonly defined as a representation of a conceptualization related to a knowledge domain, which accounts for a number of *classes* arranged hierarchically, *relations* among them, and *instances* of such classes, also called *individuals* [13]. One way to define ontologies is through *expert elicitation*, where experts in a certain knowledge domain codify their elements and relations. For example, an expert could model an office environment by defining the type of objects that usually appears in it (classes), e.g. `Table`, `Chair`, `Computer_screen`, etc., and establishing their contextual properties (relations), e.g. `Table hasOrientation Horizontal`. Relations can also set associations between classes, e.g. `Chair isNear Table`, which expresses that chairs are normally placed near tables. Knowledge about the objects from a particular scenario and their properties can be stated in the ontology through instances, e.g. `table-1`, `chair-1`, and instantiations of relations, "`table-1 isNear chair-1`". Figure 3-bottom shows part of the ontology used in this work, while figure 3-top depicts, as an illustrative example, the definition of the class `Table_top` through a number of relations using the Protégè software [19]. This software codifies the resultant ontology into the OWL language [20].

The relations that characterize a class can be seen as properties, which are useful to describe the typical shape, size or relative position of its instances. For example, the relation "`Object has_area MetricMeasurement`" is used to codify the instances of the class `Object` that have an area of $MetricMeasurment$. The subclasses of `MetricMeasurement` discretizes real values into intervals, and have the form `MM_AroudXX`, which means that the measure is in the interval of the value `XX`. However, not all the instances of a class have the same appearance in the real world. To quantify that variability, properties describing the geometry of a class are annotated into the ontology with a discrete value from the set $R_A = \{null, veryLow, low, medium, high, veryHigh\}$. For example, the definition of the class `Table_top` in figure 3-top given by an expert, encodes that tables often share a common height around $0.70m$, although their area can largely vary around their averaged value, $1m^2$.

3.2 Hypotheses Generation

One of the drawbacks of PGMs is that their complexity considerably augments when the number of objects in the scene and the candidate

[2] In this work we use Pellet [18] as logical reasoner.

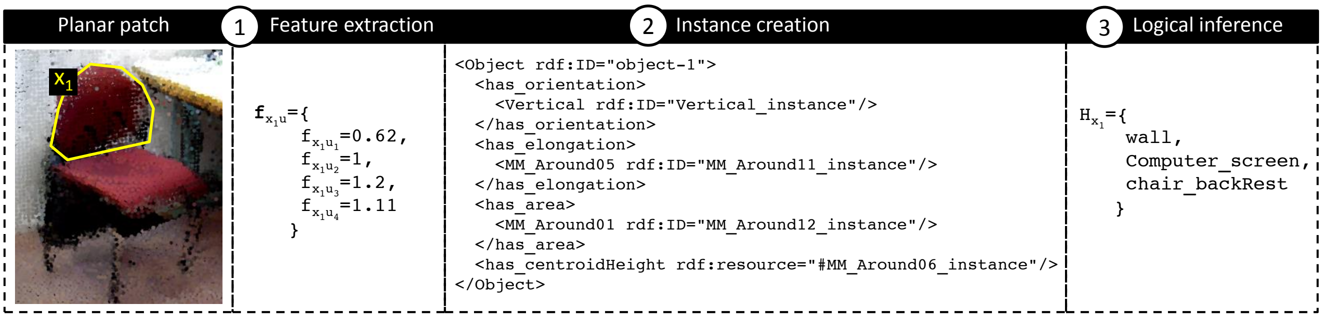

Figure 4. Example of hypotheses generation for a given planar patch. The instance is inserted into the ontology using the OWL language.

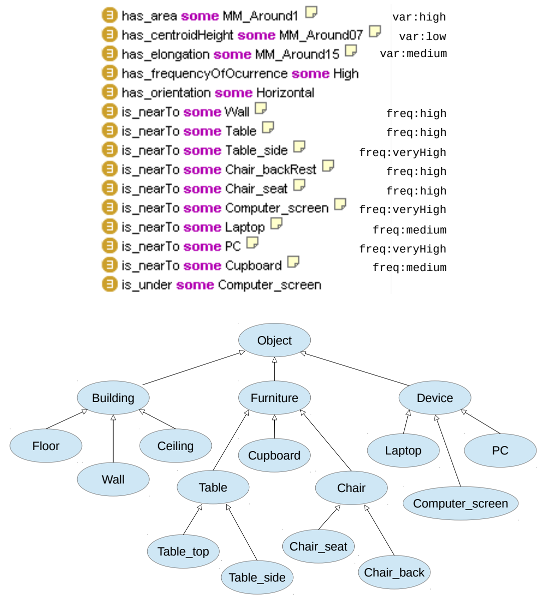

Figure 3. Top, definition of the class `Table_top`. Bottom, part of the used ontology defined by expert elicitation.

classes are large, which can produce a degradation in the quality of the recognition results. Semantic information is used in this work to mitigate this effect by hypothesizing about the potential classes of the observed objects. The hypotheses generation process is as follows.

Given the set $L = \{l_1, \dots, l_k\}$ of the considered k classes of the domain, and a planar patch x_i to be recognized, a new individual derived from the `Object` class is created into the ontology, e.g. `object-1`, annotating its unary features $f(x_i) = \boldsymbol{f}_{x_i u_i}$ through the same relations shown in figure 3-top. For example, if a patch has a centroid height of 0.73 meters from the floor, the relation "`object-1 hasCentroidHeight MM_Around07`" is added to the ontology. Once the instance is properly characterized, a logical reasoner, Pellet [18] in our implementation, infers a set of classes, $H_{x_i} \subseteq L$, which include such a relation in their definitions.

This process is performed for all the n observed objects, obtaining a set of hypothetical classes $H = \{H_{x_1}, H_{x_2}, ..., H_{x_n}\}$ that, as it is shown in the evaluation section, is drastically smaller than the initial one. Figure 4 shows an example of hypotheses generation for a `Chair_backRest`.

During the hypotheses generation process and in order to cope with the variability that objects may exhibit in the scene, we rely on the annotations over the relations defined in the ontology[3]. For example, consider the `Table_top` definition in figure 3-top that encodes a "low" height variation from the average value, i.e. $70cm$. This semantic information is used to spread out the definition widening the interval from $60cm$ to $80cm$, codified as: "`has_centroidHeight some MM_Around06 or MM_Around07 or MM_Around08`". Such a range expansion takes into account the average relation value of the different classes, i.e., the `has_centroidHeight` relation of an object class with an average of $3.6m$ suffers a higher spreading in comparison with a class with height $0.3m$, supposing that both classes are annotated with the same value from the set R_A. Note that the selection of the interval widths and the measurement discretization are provided by expert elicitation according to the particular domain at hand.

It is worth to mention that an additional advantage of using such hypotheses as class candidates is that the recognition results provided by the probabilistic reasoning over the CRF will be coherent with the information in the ontology, and consequently, with the semantic knowledge that experts encode about the domain.

3.3 Frequency of occurrence prior

Unary factors $U(\cdot)$ in a CRF give information about the compatibility of a certain object x_i w.r.t a set of classes H_i according to its appearance and geometry. This can be viewed as a way to model the probability distribution $P(y_i|x_i)$. On the other hand, pairwise factors codify $P(y_i, y_j|x_i, x_j)$, i.e., how probable a class assignation for two objects becomes given their relational (contextual) features. Thus, by combining both factors, the CRF can exploit appearance, geometric and contextual features. In this section we propose the addition of prior information about the frequency of occurrence of objects to the CRF formulation, which can help to disambiguate some recognition results. Prior information is added to the unary factors, which now encode the product of two probabilities, i.e:

$$U(y_i, x_i, \boldsymbol{\omega}) \approx P(y_i|x_i, \boldsymbol{\omega})P(y_i) \tag{6}$$

Prior information is codified into the ontology through the relation `has_frequencyOfOccurrence`, which takes values from the set R_A. In order to adapt the probability distribution

[3] Notice that these annotations could have been introduced as additional relations, e.g., `has_area_variability`. However, given that the logic reasoner is not going to take advantage of them, and aiming to have a representation as clear as possible, we decided to use annotations.

Figure 5. Mobile robot Rhodon gathering 3D data from an office.

$P(y_i)$ to the linear classification model shown in equation 1, it is replaced by the function $f_o(y_i) : R_A \rightarrow [0..1]$, which can be considered as a non-normalized version of the former probability. For example, if the class `Chair_back` is defined in the office domain with the relation "`Chair_back has_frequencyOfOccurrence veryHigh`", and `Computer_screen` with "`Computer_screen has_frequencyOfOccurrence medium`", the f_o function can be defined to produce $f_o(Chair\_back) = 0.9$ and $f_o(Computer\_screen) = 0.5$. Thus, we define an unary factor as follows:

$$U(y_i, x_i, \boldsymbol{\omega}) = \sum_{l \in L} \delta(y_i = l)\boldsymbol{\omega}_l f(x_i) f_o(y_i) \tag{7}$$

Conversely to the hypotheses generation case, here the function $f_o(\cdot)$ is independent of the scene, so it can be computed once and stored in a look-up table, so speeding up the recognition process.

4 METHOD EVALUATION

In order to evaluate our object recognition method, we have collected a dataset compounds of 25 office scenes using the mobile robot Rhodon, which is endowed with a Kinect-like sensor mounted on a Pan-Tilt unit (see figure 5).

The planar patches extracted from the scene fed a CRF, which is trained using synthetic data as explained in [14], and are also employed to generate hypotheses about their most probable classes, as described in section 3.2. In our experiments we have considered a total of 7 object classes: L={`Floor`, `Wall`, `Table_top`, `Table_side`, `Chair_bakRest`, `Chair_seat`, `Computer_screen`}.

The performance of our approach has been measured using the micro/macro precision/recall metrics [2]. Table 2 shows the results obtained using 4 different recognition variants on the 25 considered scenarios. The first variant only uses appearance and geometric object features, achieving a micro p./r. of $\sim 79\%$, while the second one exploits contextual relations and increases that percentage by $\sim 5\%$. The third variant incorporates the generation of hypotheses, reaching a micro p./r. of $\sim 93.5\%$, and the last one also uses prior information, obtaining $\sim 94.3\%$ of success. These results proves that contextual information improves the recognition of objects in a scene, and that the use of semantic information prevents the CRF from providing non-coherent results, increasing thus the recognition success. Prior information also adds a sense of coherence to the method by yielding the frequency of occurrence of the different object classes in an office environment, which is reflected as an improvement in the results.

Table 2. Results of the tests conducted

Variant used	micro p./r.	macro p.	macro r.
(1) No context	79.23	78.35	77.52
(2) Context	84.07	86.11	87.68
(3) Context + Hypotheses	93.45	92.52	92.45
(4) Context + Hypotheses + Prior	94.31	93.69	93.28

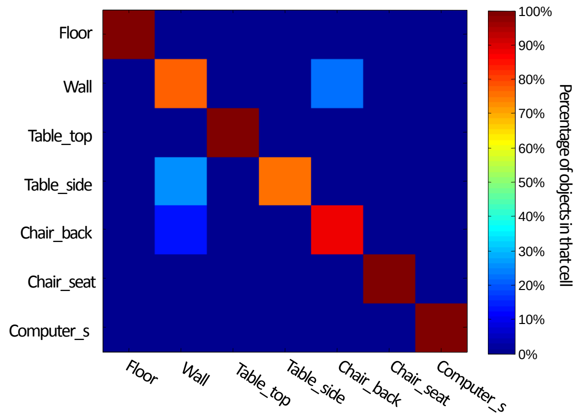

Figure 6. Confusion matrix of the actual object classes and the recognition results.

Figure 6 shows the confusion matrix obtained using the last variant, where rows represent the actual class of the objects, and columns the class which they are recognized as. We can see how erroneous recognitions correspond to the classes `Wall`, `Table_side` and `Chair_back`, given that, with the considered features, it is sometimes difficult to differentiate them.

Part of the achieved improvement in the performance using the generation of hypotheses is due to the fact that it enables *probabilistic exact inference*, i.e., the checking of all the possible classes assignations for the scene objects. To illustrate that, let's consider the scenario shown in figure 7-top, entailing 11 objects. Given that we have considered 7 object classes, probabilistic reasoning by exact inference consists in computing equation 3 a total of 7^{11} times. Such a computation takes several hours, which is unfeasible for a mobile robot that requires quick results in order to operate within real environments. However, relying on the generated hypotheses as candidate classes, the number of combinations is reduced, in this example, to 1536, which can be computed in a few milliseconds. Figure 7-bottom shows the objects from 7-top recognized through an exact inference process.

In addition, the robot can use the probability associated to the recognition results as a measure of uncertainty. Thus, results with high uncertainty could motivate the execution of further actions by the robot, like gathering additional data from the scene, in order to obtain more plausible results.

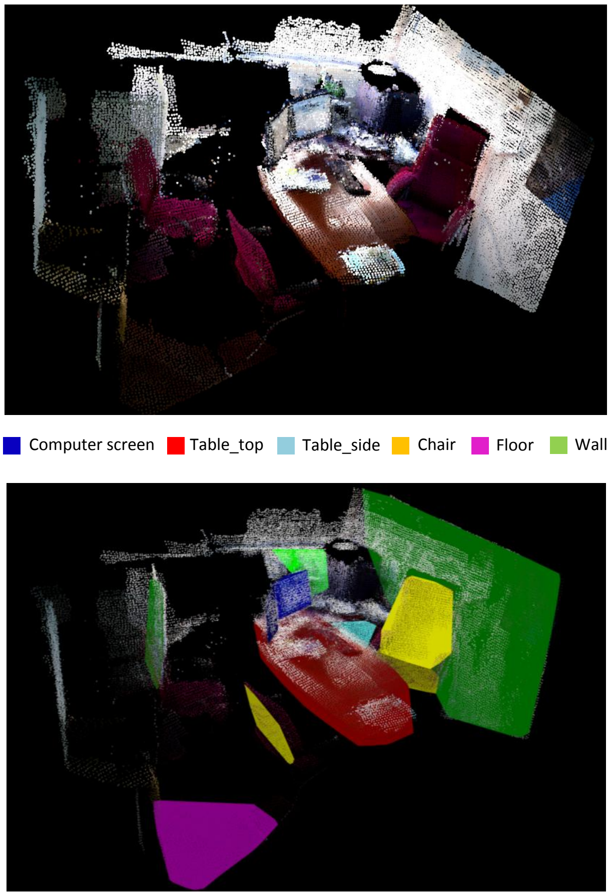

Figure 7. Recognition result in one of the studied scenarios. Top, 3D data from an office environment. Bottom, planar patches detected and recognized using our approach.

5 CONCLUSIONS

In this work we have addressed a key problem that mobile robots intended to operate in human environments have to solve: the recognition of objects within the scene. For that, we have proposed an approach that makes a combined use of Probabilistic Graphical Models (PGMs) and Semantic Knowledge (SK). PGMs, which can handle uncertainty, are used to recognize objects using a probabilistic reasoning, while SK is used to: hypothesize about the most promised class candidates for objects (reducing in this way the complexity of the PGMs), provide prior information about the frequency of occurrence of the different object classes (gaining in robustness and accuracy in the recognition results), and maintain a representation of the environment that enables high level robotic tasks. We have conducted a number of experiments that validate our approach, yielding a recognition success of $\sim 94\%$.

In the future, we plan to study how to deal with objects showing unusual properties. Let's suppose a scene with a computer screen placed on the floor. In that situation the logical reasoner does not yield the class `Computer_screen` as a hypothesis, given that its height largely differs from the expected one. An option could be to consider the result of the logical inference as a score in the CRF formulation, at the cost of removing the *exact inference* option.

ACKNOWLEDGEMENTS

We want to express our gratitude to our colleague E. Fernandez-Moral for both, providing us the implementation of the plane-based mapping algorithm, and his support during the collection of the office dataset employed in the method evaluation. This work has been funded by the Spanish grant program FPU-MICINN 2010 and the Spanish project "TAROTH: New developments toward a robot at home".

REFERENCES

[1] X. Xiong, D. Huber, Using Context to Create Semantic 3D, in Proceedings of the British Machine Vision Conference (BMVC), 2010.
[2] A. Anand , H. S. Koppula , T. Joachims, A. Saxena, Contextually Guided Semantic Labeling and Search for 3D Point Clouds, in The Int. Journal of Robotics Research January vol. 32 no. 1 19-34, 2013.
[3] Y. Xiang, X. Zhou, Z. Liu, T. Chua, C. Ngo, Semantic Context Modeling with Maximal Margin Conditional Random Fields for Automatic Image Annotation, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010.
[4] Y. Duan, C. Cruz, and C. Nicolle, Architectural Reconstruction of 3D Building Objects through Semantic Knowledge Management, in SNPD, 2010.
[5] C. Galleguillos, S. Belongie, Context based object categorization: A critical survey, in Computer Vision and Image Understanding vol. 114, n. 6, 2010.
[6] D. Koller, N. Friedman, Probabilistic graphical models: principles and techniques, MIT press, 2009.
[7] A. Quattoni, M. Collins, T. Darrell, Conditional random fields for object recognition, in Neural Information Processing Systems, 2004.
[8] J. Besag, On the statistical analysis of dirty pictures, Journal of Royal Statistical Society. Series B (Methodological), 48(3) 259302, 1986.
[9] Yuri Boykov, Olga Veksler, Ramin Zabih. Fast Approximate Energy Minimization via Graph Cuts. In IEEE transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 23, no. 11, pp. 1222-1239, 2001.
[10] Kevin P. Murphy, Yair Weiss, and Michael I. Jordan. Loopy belief propagation for approximate inference: an empirical study. In Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence (UAI), 1999.
[11] H. B. Hmida, C. Cruz, F. Boochs, C. Nicolle, Knowledge Base Approach for 3D Objects Detection in Point Clouds Using 3D Processing and Specialists Knowledge, in International Journal On Advances in Intelligent Systems, 2013.
[12] C. Galindo, J.A. Fernandez-Madrigal, J. Gonzalez-Jimenez, A. Saffiotti, Robot Task Planning using Semantic Maps, in Robotics and Autonomous Systems, 2008.
[13] M. Uschold, M. Gruninger, Ontologies: principles, methods and applications, in The Knowledge Engineering Review, 11, pp 93-136, 1996.
[14] J.R. Ruiz-Sarmiento, C. Galindo, J. Gonzalez-Jimenez, Exploiting Semantic Knowledge for Robot Object Recognition, in Artificial Intelligence Journal (submitted), 2013.
[15] C. Galindo, A. Saffiotti, Inferring Robot Goals from Violations of Semantic Knowledge, in Robotics and Autonomous Systems, 2013.
[16] C. Galindo, J-A Fernandez-Madrigal, Javier González. Multihierarchical interactive task planning: application to mobile robotics. In IEEE Transactions on Cybernetics, vol. 38. pp 785-798, 2008.
[17] E. Fernandez-Moral, W. Mayol-Cuevas, V. Arevalo, J. Gonzalez-Jimenez, Fast place recognition with plane-based maps, IEEE International Conference on Robotics and Automation (ICRA), 2013.
[18] E. Sirin, B. Parsia, B C. Grau, A. Kalyanpur and Y. Katz, Pellet: A practical owl-dl reasoner, in Web Semantics: science, services and agents on the World Wide Web, 2007.
[19] Protege, Home page, http://protege.stanford.edu/.
[20] M. Dean, G. Schreiber, OWL web ontology language reference, http://www.w3.org/TR/2004/REC-owl-ref-20040210/ (2004).