Knowledge-Based Systems 86 (2015) 131–142

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

# Exploiting semantic knowledge for robot object recognition

José-Raúl Ruiz-Sarmiento *, Cipriano Galindo, Javier Gonzalez-Jimenez

*System Engineering and Automation Dept., University of Málaga, Campus de Teatinos, 29071 Málaga, Spain*

ARTICLE INFO

*Article history:*
Received 7 January 2015
Received in revised form 13 May 2015
Accepted 27 May 2015
Available online 8 June 2015

*Keywords:*
Semantic knowledge
Human elicitation
Object recognition
Probabilistic Graphical Models
Autonomous robots

ABSTRACT

This paper presents a novel approach that exploits semantic knowledge to enhance the object recognition capability of autonomous robots. Semantic knowledge is a rich source of information, naturally gathered from humans (elicitation), which can encode both objects' geometrical/appearance properties and contextual relations. This kind of information can be exploited in a variety of robotics skills, especially for robots performing in human environments. In this paper we propose the use of semantic knowledge to eliminate the need of collecting large datasets for the training stages required in typical recognition approaches. Concretely, semantic knowledge encoded in an ontology is used to synthetically and effortless generate an arbitrary number of training samples for tuning Probabilistic Graphical Models (PGMs). We then employ these PGMs to classify patches extracted from 3D point clouds gathered from office environments within the UMA-offices dataset, achieving a ∼90% of recognition success, and from office and home scenes within the NYU2 dataset, yielding a success of ∼81% and ∼69.5% respectively. Additionally, a comparison with state-of-the-art recognition methods also based on graphical models has been carried out, revealing that our semantic-based training approach can compete with, and even outperform, those trained with a considerable number of real samples.



## 1. Introduction

Object recognition is one of the key abilities of a mobile robot intended to perform high-level tasks in human environments, where objects are usually placed according to their functionality, e.g., tv-sets are in front of couches, night tables are near beds, etc. As reported by other authors [11], the exploitation of these contextual relations, that can be seen as a form of *semantic knowledge*, can improve the performance of traditional object recognition methods which only rely on sensorial features.

To illustrate the benefits of using semantics, let us consider a robot coping with the task of recognizing the objects placed in its surroundings. This may become complex for a number of reasons, including the large number of possible object classes and features to extract, their similarity, etc. Suppose now that the robot knows that it is in an office and has some semantic knowledge related to that particular domain, for example the type of objects usually present in a typical office environment and their contextual relations. This simplifies the recognition problem, drastically reducing the range of possible objects classes, and even more importantly, enabling the recognition system to exploit particular object relations to gain in effectiveness and robustness. For instance, an object that resembles an office table according to its geometry can be more confidently recognized as such if objects typically found near it, e.g. a computer screen and/or a chair, are also detected and fulfill certain contextual relations, for example, the computer screen is on the table and the chair is close to it.

In this work we present a novel approach that exploits semantic knowledge encoded by *human elicitation* to train *Probabilistic Graphical Models* (PGMs) [16] for object recognition. PGMs form a machine learning framework that is widely applied to object recognition given its capabilities for modelling both uncertainty and objects relations. These systems need a vast amount of training data in order to reliably encode the gist of the domain at hand, however, the gathering of that information is an arduous, time-consuming, and – in some domains – not a tractable task. To face this issue, we codify semantic knowledge by means of an ontology [30], which defines the domain object classes, their properties, and their relations, and use it to generate training samples for a Conditional Random Field (CRF) [16]. These training samples reify prototypal scenarios where objects are represented by a set of geometric primitives, e.g., planar patches or bounding boxes, that fulfill certain geometric properties and relations, like proximity, difference of orientation, etc.

Aiming to show the performance of CRFs trained with the proposed approach, they have been integrated into an object

* Corresponding author.
*E-mail addresses:* jotaraul@uma.es (J.-R. Ruiz-Sarmiento), cipriano@ctima.uma.es (C. Galindo), javiergonzalez@uma.es (J. Gonzalez-Jimenez).

http://dx.doi.org/10.1016/j.knosys.2015.05.032
0950-7051/

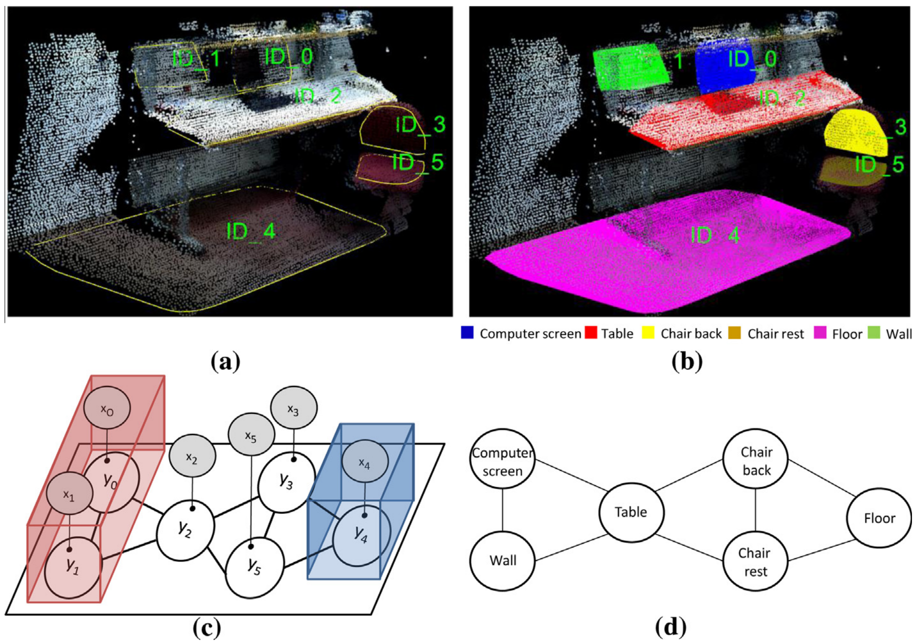


**Fig. 1.** (a) Example of a scene segmented into planar patches (labeled with an ID and delimited by yellow lines). (b) Scene objects recognized by our method. (c) Graphical model built for the planar patches shown in (a). Each patch is associated to a node $y_0, \ldots, y_5$, whose probabilistic distributions are conditioned to their respective patch observations $x_0, \ldots, x_5$ (observation $x_i$ corresponds to patch *ID_i*). Near patches are linked by an edge. The blue box encapsules the scope of a particular unary factor, while the red one shows the scope of a pairwise factor. (d) The resultant graphical model after the execution of the recognition method, when random variables take a value according to their most probable assignation.

recognition framework. This framework operates by processing point clouds provided by a RGB-D camera, in order to extract geometric primitives (see Fig. 1(a)), which are then recognized as belonging to a certain object class through an inference process over the trained CRF. We have obtained promising results in office and home environments, employing both planar patches and bounding boxes as geometric primitives, though our methodology can be applied to other scenarios and sensorial data types.

In the literature, PGMs are used, in general, to learn the properties of the different object classes and their contextual relations using data from previously collected datasets. In contrast, the work presented here drives this learning phase by providing synthetic training samples extracted from the semantic knowledge of the domain at hand. This knowledge can be naturally provided by humans and encoded into an ontology, and exhibits three advantages with respect to other related approaches:

- It eliminates the usually complex and high resource-consuming task of collecting the large number of training samples required to tune an accurate and comprehensive model of the domain.
- Ontologies are compact and human-readable knowledge representations. In that way, extending the problem with additional object classes is just reduced to codify the knowledge about the new classes into the ontology, generate synthetic samples considering the updated semantic information, and train the CRF. This process can be completed in a few minutes, in contrast to the time needed for gathering and processing real data.
- The recognized objects are anchored to semantically defined concepts, which is useful for robot high-level tasks like reasoning or task planning [10,8,4].

We have conducted an evaluation of our work employing two datasets: one from our facilities, called UMA-offices, which counts 25 office environments, and the NYU2 dataset [28], from which we have extracted 61 offices and 200 home scenes. The performance of CRFs trained with our methodology has been also compared with two state-of-the-art methods, namely (i) a standard formulation of CRFs trained and tested with real data [16], and (ii) the CRF presented in Xiong and Huber [34]. The results show that our approach can compete with, and even outperform, those trained with a considerable number of real samples.

In the next section we put our proposal in the context of other related works. Section 3 introduces Probabilistic Graphical Models applied to object recognition, while in Section 4 we present the proposed method to train these models using semantic knowledge. In Section 5, the evaluation results of the method considering two datasets comprising office and home environments are shown, and a comparison with other state-of-the-art approaches is presented. Finally, Section 6 ends with some conclusions and future work.

## 2. Related work

Object recognition is a key topic in robotics and computer vision that, in many cases, has been successfully addressed by *only* using the visual features of isolated objects, i.e. without considering information from the rest of the scene. Some remarkable examples are the Viola and Jones boosted cascade of classifiers [32], the SIFT object recognition algorithm [19] or the Bag of Features [21] models. However, the current trend also considers the exploitation of contextual information between objects, aiming to improve the recognition results (see Galleguillos and Belongie [11]).

Throughout this section, we discuss related works on object recognition systems that resort to graphical models or semantic knowledge to model contextual information. Also, some works reporting different alternatives to the utilization of ontologies as a source of semantic information for object recognition are commented.

### 2.1. Probabilistic Graphical Models

Probabilistic Graphical Models (PGMs) [16] is one of the most resorted frameworks to manage contextual information. The