Probability and Common-Sense: Tandem Towards Robust Robotic Object Recognition in Ambient Assisted Living

J.R. Ruiz-Sarmiento, C. Galindo, and J. Gonzalez-Jimenez

Machine Perception and Intelligent Robotics Group, System Engineering and Automation Dept., University of Málaga, Campus de Teatinos, 29071, Málaga, Spain, {jotaraul,cgalindo,jgonzalez}@uma.es

Abstract. The suitable operation of mobile robots when providing Ambient Assisted Living (AAL) services calls for robust object recognition capabilities. Probabilistic Graphical Models (PGMs) have become the de-facto choice in recognition systems aiming to efficiently exploit contextual relations among objects, also dealing with the uncertainty inherent to the robot workspace. However, these models can perform in an incoherent way when operating in a long-term fashion out of the laboratory, e.g. while recognizing objects in peculiar configurations or belonging to new types. In this work we propose a recognition system that resorts to PGMs and common-sense knowledge, represented in the form of an ontology, to detect those inconsistencies and learn from them. The utilization of the ontology carries additional advantages, e.g. the possibility to verbalize the robot's knowledge. A primary demonstration of the system capabilities has been carried out with very promising results.

Keywords: AAL, robotics, object recognition, probabilistic graphical models, ontologies, environmental information

1 Introduction

Ambient Assisted Living (AAL) aims to facilitate and extend the independence of elderly people by means of cutting-edge solutions from Information and Communication Technologies (ICT). Mobile robots have stood out as valuable ICT systems to improve the quality of life of the elderly, being applied with success to a variety of services: health care, companion, entertainment, household maintenance, etc [3]. A required robotic ability to provide those services with guarantees is such of object recognition. Moreover, the robot should exhibit the capability to reason about what it is perceiving and, if needed, to react in consequence. For example, if the robot recognizes a pill box on a counter, and it knows that its content is perishable and needs refrigeration, it should infer the need to

Work supported by the research projects TEP2012-530 and DPI2014-55826-R, funded by the Andalusia Regional Government and the Spanish Government, respectively, both financed by European Regional Development's funds (FEDER).

put it into the fridge or alert the elder – this option also requires Human-Robot Interaction (HRI) capacities.

A renowned tool to tackle the robotic object recognition problem is such of Probabilistic Graphical Models (PGMs), and concretely Conditional Random Fields (CRFs) [2]. These graph-based models provide a mathematically-grounded framework for the recognition of objects exploiting their contextual relations, also dealing with uncertainty. Home environments are rich in contextual information, which is useful to disambiguate recognition results, e.g. a white box near a fridge is more probable to be a microwave than a night-stand [5]. Typically, mobile robots employ CRFs that are pre-tuned with a certain dataset in order to recognize a fixed range of object categories. However, this configuration lacks of the flexibility demanded by robots performing in home environments, e.g. it is (of course) unable to recognize new types of objects not appearing in the training dataset, or instances of learned ones showing peculiar features, which can lead to an incoherent performance.

In this paper we propose a CRF-based object recognition system that relies on common-sense knowledge about home environments to detect and learn from incoherent recognition results. Concretely, common-sense knowledge is codified as an ontology [8], which allows for a natural definition of the properties of concepts in the domain (object types in this case) and their relations. For example, we can define the concept `Fridge` codifying that they are usually tall, box-shaped objects, and the `Pill_box` one stating that they are small boxes related to fridges by `Pill_box placedInto Fridge`. In the proposed system, the recognition results yielded by probabilistic inference over the CRF are checked for coherence against the common-sense representation. If any of them is detected as incoherent (for example, a middle-size object is classified as a fridge), then it is annotated for its posterior evaluation by the user through a simple dialogue. This human-robot interaction is greatly supported by the ontology, since its content can be verbalized in a straightforward way. Finally, the feedback from the user is back-propagated in order to tune the CRF accordingly (see Fig. 1 for a complete overview of the system). It is worth to mention that ontologies also suppose a basic way to *understand* the robot workspace, enabling the detection of object configurations that can be hazardous, e.g. the pill box found out of the fridge.

Summarizing, the presented object recognition system is able to: (i) exploit contextual relations, (ii) handle uncertainty, (iii) detect incoherent results, (iv) learn from experience, and (v) verbalize its outcome. In this paper we introduce the ongoing research towards capabilities (iii) and (iv). To illustrate the system capabilities we have conducted a proof-of-concept demonstration showing promising results.

2 Proposed Robotic Object Recognition System

This section, first, gives some brief background on the theoretical concepts that support our recognition system, and ends up with a description of its operation.

2.1 Conditional Random Fields for Object Recognition

In order to recognize the objects appearing in a given scene, Conditional Random Fields (CRFs) [2] build a graph representation of it according to the objects' layout. In this graph, usually denoted by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the nodes $\mathcal{V}$ represent random variables, and the edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ link nodes that are related in some way. The random variables, commonly represented as $\boldsymbol{y} = [y_1, ..., y_n]$, take values from a set $\mathcal{L}$ of considered object types, e.g. chair, table, tv, fridge, etc., and are associated with the vector of n object observations $\boldsymbol{x} = [x_1, ..., x_n]$. In this way, only the nodes representing objects that are close to each other in the scene are linked with an edge (see Fig. 1-a and 1-b). CRFs efficiently encode the probability distribution $p(\boldsymbol{y}|\boldsymbol{x})$ over this graph representation by:

$$p(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{x},\boldsymbol{\theta})} \prod_{c \in C} \exp(\langle \phi(x_c, y_c), \boldsymbol{\theta} \rangle) \tag{1}$$

where $\mathcal{C}$ is the set of maximal cliques of the graph $\mathcal{G}$ (in this work we consider cliques of size one, nodes, and two, neighboring nodes), $Z(\cdot)$ is the also called partition function, which plays a normalization role so $\sum_{\xi(\boldsymbol{y})} p(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}) = 1$, being $\xi(\boldsymbol{y})$ a possible assignment to the variables in $\boldsymbol{y}$. $\langle \cdot, \cdot \rangle$ stands for the inner product, and $\phi(x_c, y_c)$ are the sufficient statistics of the variables in the clique c. These sufficient statistics comprise the salient features of the data, e.g. features characterizing nodes in $\mathcal{V}$ are the height of the object, orientation, color, etc., while features about the relations in $\mathcal{E}$ are difference in height, difference of orientation, etc. The vector $\boldsymbol{\theta}$ stands for the model parameters (or weights) to be tuned during the CRF training.

Once the scene and the objects observed therein have been represented as a graph, a probabilistic inference algorithm can be executed over it in order to retrieve the recognition results. This is usually carried out by algorithms computing the Maximum a Posteriori (MAP) assignment [2], i.e. the assignment $\hat{\boldsymbol{y}}$ to the variables in $\boldsymbol{y}$ that maximizes Eq. 1 (see Fig. 1-c).

2.2 Common-sense Knowledge Representation

Ontologies are widely-resorted representations for codifying common-sense knowledge about a certain domain. They are often created by an expert in that domain, and consist of: a number of descriptions of concepts arranged hierarchically, relations among them, and instances of those concepts [8]. For example, concepts codifying the types of objects often appearing in home environments are `Fridge`, `Cereal_box`, or `Cabinet`, with properties like `Fridge has_orientation Vertical` or `Cereal_box has_shape Box`. Relations set associations between concepts, e.g. `Cereal_box isNear Cabinet`, which expresses that cereal boxes can be found near cabinets. Knowledge about the objects from a particular scenario and their properties can be stated in the ontology through instances, e.g. `cereal_box-1`, `cabinet-1`, and instantiations of relations, `cereal_box-1 is_near cabinet-1`. Fig. 1-d) shows the simplified ontology used in this work

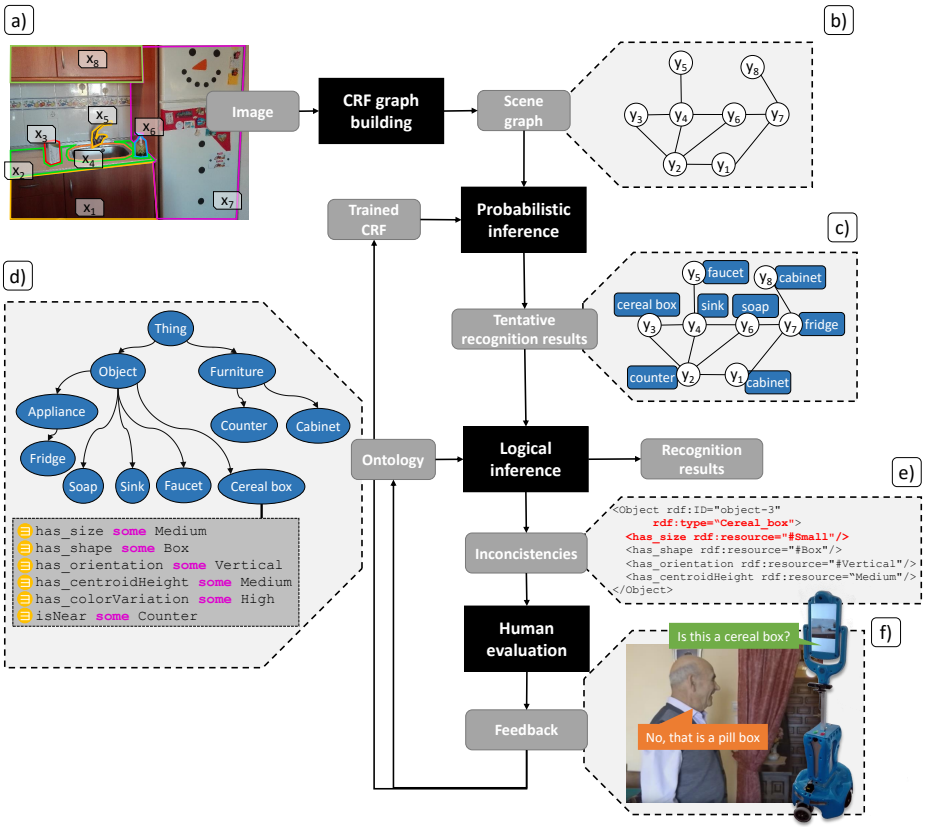

Fig. 1. Overview of the proposed object recognition system. Information about the sub-figures from a) to f) is detailed in the text.

describing objects typically appearing in kitchens, and the detail of the description of the concept `Cereal_box`.

2.3 System Overview

An overview of the complete system is depicted in Fig. 1. The recognition pipeline starts by capturing an image of the scene to be processed, and the posterior building of its CRF graph representation. This graph, along with the pre-trained CRF parameters, is exploited by a probabilistic inference algorithm to provide a set of tentative object recognition results. These results are then inserted as instances in the Ontology, which checks their consistency with respect to the codified common-sense knowledge by employing a logical reasoner. This permits the robot to detect incoherent results that are subsequently evaluated by the user. The evaluation of a conflicting object starts by showing him/her a cropped image of it. Three different scenarios are then possible:

Case 1: the user determines that the recognition result is right. This means that the CRF performed correctly, but the codified common-sense knowledge was somehow *too strict*. The ontology learns from this outcome by relaxing the codified object property that produced the inconsistency.

Case 2: the recognition result is wrong, and:

- **Case 2.1:** the object type is already present in the CRF/ontology. In this case the CRF misclassified the object. To learn from the mistake, the gathered object information is used to re-tune the CRF parameters.
- **Case 2.2:** the object type is new. The relevant information from the object is used to automatically generate a new concept in the ontology, and the CRF is also re-trained taking into account this new object type.

The re-training of a CRF model entails the utilization of a dataset with a considerable number of samples covering all the relevant object types. Since this dataset consist of the sufficient statistics extracted from objects and relations, and raw data are not required (e.g. images), its size in memory is usually small. However, if the robot storage is quite limited, we have shown in a previous work [4] how an ontology can be also exploited to successfully train CRFs.

3 System Demonstration

To carry out a primary demonstration of the system capabilities we trained a CRF to recognize the same object types that were codified into the ontology. This training was carried out through the UPGMpp library [6], relying on the pseudo-likelihood objective function and the L-BFGS optimization method. Regarding the ontology design, we resorted to Protégè[2]. The CRF and the ontology were then plugged into the recognition system as shown in Fig. 1, which was integrated into the mobile robot Giraff [1] (see Fig. 1-f). This robot is built upon a motorized wheeled platform endowed with a 2D laser scanner, a RGB-D camera, and a videoconferencing set: microphone, speaker, and screen.

The robot was deployed into an apartment and performed a primary task: to check the configuration of the objects in the kitchen. Concretely, during the robot operation, the RGB-D camera was used to capture both intensity and depth images when reaching certain locations in the kitchen. For brevity we analyze one system trace reported while detecting an inconsistency due to a new object category. This is the most complete case of the three aforementioned ones.

The described trace is fully depicted in Fig.1. It starts with the robot capturing the image shown in Fig. 1-a, and the recognition system building its corresponding CRF-graph (Fig. 1-b). After a probabilistic inference process [2], the MAP assignment yields the tentative object recognition results (Fig. 1-c). These results are inserted as instances of objects in the ontology, which checks their consistency with respect to the encoded knowledge. The logical inference process [7] detects an inconsistency: the object x_3 was recognized as a cereal box, but it does not exhibit the common size of cereal boxes (Fig. 1-e, expressed

[2] http://protege.stanford.edu/

as OWL). The remaining results are considered as valid, while the inconsistent one has to be evaluated by the user. Through a simple dialogue (Fig. 1-f), the user determines that the object (x_3) is in fact a pill box, and that it must be stored in the fridge. This information is then back-propagated to both: (i) the ontology, where the system creates a new concept `Pill_box`, inheriting from the `Object` one, and describes it with the information gathered from the human and from the collected sensory data, and (ii) to the CRF model, which re-tunes its parameters according to the new information. The learning success was evaluated in later observations of pill boxes, where the robot was able to successfully recognize this new type of object.

4 Discussion and Future Work

In this work we have proposed a robust robotic object recognition system for providing AAL services consisting of two parts: a Conditional Random Field (CRF), and an ontology. The system is able to detect incoherent recognition results and learn from them including a human in the loop, aiming to improve its performance and robustness in the long-term operation within home environments. A demonstration of its suitability has been carried out, showing promising results.

There is significant room to explore possible system improvements or applications, including a thorough evaluation of the system with more complex CRFs and ontologies. On the other hand, both CRFs and ontologies could be extended to also consider information relative to rooms, opening new possibilities to exploit object-room relations, e.g. that fridges are usually in kitchens. The system could also benefit from a study of when would be more appropriate to ask the user about inconsistent results in order to not bother him/her.

References

1. Gonzalez-Jimenez, J., Galindo, C., Ruiz-Sarmiento, J.R.: Technical improvements of the giraff telepresence robot based on users' evaluation. In: RO-MAN (2012)
2. Koller, D., Friedman, N.: Probabilistic Graphical Models: Principles and Techniques. MIT Press (2009)
3. Payr, S., Werner, F., Werner, K.: Potential of robotics for ambient assisted living. Tech. rep., Austrian Research Institute for Artificial Intelligence (2015)
4. Ruiz-Sarmiento, J.R., Galindo, C., González-Jiménez, J.: Exploiting semantic knowledge for robot object recognition. Know.-Based Systems 86, 131–142 (2015)
5. Ruiz-Sarmiento, J.R., Galindo, C., González-Jiménez, J.: Joint categorization of objects and rooms for mobile robots. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2015)
6. Ruiz-Sarmiento, J.R., Galindo, C., González-Jiménez, J.: UPGMpp: a Software Library for Contextual Object Recognition. In: 3rd. REACTS Workshop (2015)
7. Sirin, E., Parsia, B., Grau, B.C., Kalyanpur, A., Katz, Y.: Pellet: A practical owl-dl reasoner. Web Semantics 5(2), 51–53 (Jun 2007)
8. Uschold, M., Gruninger, M.: Ontologies: principles, methods and applications. The Knowledge Engineering Review 11, 93–136 (1996)