Autonomous Object Manipulation: A Semantic-Driven Approach*

Nicola Vitucci
Dipartimento di Elettronica e Informazione (DEI)
Politecnico di Milano
Milano, Italy
vitucci@elet.polimi.it

Abstract

The problem of grasping is widely studied in the robotics community. This project focuses on the identification of object graspable features using images and object structural information. The primary aim is the creation of a framework in which the information gathered by the vision system can be integrated with automatically generated knowledge, modelled by means of fuzzy description logics.

1 Introduction

Research on grasping and manipulation is nowadays focusing on how to make this task as autonomous as possible: recent advances in the fields of computer vision, knowledge engineering and neurophysiology show that a variety of techniques can be used to get closer to this goal. One of the major issues is given by the high dimensionality of the problem, as many variables occur both in the object recognition and in the grasp decision processes.

Different data representations have been studied to deal with the grasping problem [Rubio *et al.*, 2010; Huebner *et al.*, 2009], the main aim being to find a suitable representation of object and grasp types, in order to reduce the complexity of the problem. Most of the proposed representations focus on the visual characteristics of the objects more than on their "conceptual" representations. In our project we are exploring a semantic representation of objects using a suitable description logic (DL) formalism, which should take into account the qualitative and quantitative characteristics that can be useful for grasping and manipulation tasks.

The decomposition of an object into parts has been shown to be both close to the way human beings reason about objects [Biederman, 1987] and effective in reducing the complexity of computing tasks like object recognition (e.g., [Agarwal *et al.*, 2004]). The semantic description of the parts composing an object and their mutual relationships is a difficult problem: apart from the problems related to the image analysis phase, the vagueness affecting real world data makes formal representation of an object structure quite inconvenient. Novel tools such as fuzzy DLs [Lukasiewicz and Straccia, 2008], which extend crisp DLs by adding imprecision and vagueness in the reasoning process, may thus be used to bridge qualitative and quantitative reasoning in a unified framework.

The problem of semantic part decomposition is still an open problem and, to the best of our knowledge, there are no tools available to automatically create a fuzzy ontology from raw data taken from an image. The use of fuzzy DLs for object recognition has been investigated in some works such as [Hudelot *et al.*, 2008], in which little advantage is taken from the (partial) fuzzy extension and from the expressivity of the used logic (i.e., no cardinality restrictions are used); furthermore, a preliminary phase of semantic annotation of the images by domain experts has to be performed.

Some recent works make a thorough use of fuzzy DLs to reason about multimedia information (e.g., [Simou *et al.*, 2007]) but, again, little advantage is taken from the expressivity given by cardinality restrictions (when available). Generally speaking this is due to the fact that, for scene understanding purposes, it is sufficient to know whether an object is present or not, while, for object recognition purposes, it is often necessary to be able to count the number of similar instances of the same class.

2 Current work

As shown in the previous section, the first problem arising in an autonomous manipulation task is given by object localization and recognition. The input image has to be segmented in a meaningful way, in order to distinguish each object from the background and from other objects; we are now focusing only on the shape of the objects (thus using synthetic models), as the first step is to make the system able to build its own representation of each object. The problems due to occlusions and cluttering have to be faced after some knowledge about the object to recognize has been gathered, so to use it for trying to solve ambiguities.

The most delicate part of the process is given by the decomposition of an object in its composing parts, firstly as an image or mesh processing task. We are currently taking advantage of some concepts pertaining to the *Gestalt theory* and its application to computer vision [Desolneux *et al.*, 2004], such as proximity, closure and similarity, which can be described in a formal and efficient way. Some studies in the field of developmental robotics have been performed on how to generalize such relations by using *perceptual functions* [Modayil and Kuipers, 2007], but this is at the moment out of our scope.

*Research project supervised by Prof. Giuseppina Gini.

The formalization is built on top of fuzzy DLs, which give a chance both to ground qualitative concepts as "close to" and "connected to" to quantitative real-world data and to define a degree of membership of an object to a certain class. The available tools (as the fuzzy reasoners *FuzzyDL*[1] and *FiRE*[2]) are still under development, so there are several limitations to take into account in the formalization process.

Fuzzy DLs are used to represent the structure of the objects, e.g., "a fork is composed of a long and thin part (the handle) and four short and thin parts (the tines)", while the ontology of grasping types currently contains axioms like

$$\textsc{GraspableByPinch} \equiv \textsc{MediumLengthObject} \sqcap \textsc{HighlyEccentricalObject}$$

which are related to the geometrical characteristics of the objects (and object parts) involved. The role of the latter ontology is to determine (even offline) whether a specific kind of grasp can be performed or not on a selected object or on one of its composing parts.

From the semantic point of view, the most difficult problem is to define what is to be called a *part*; this is a well known problem in logic faced in the *mereotopology* framework [Varzi, 2007; Keet and Artale, 2008], in which formal properties of the part-whole relation are extended to take into account the notion of connection. The work on fuzzy spatial relations which has been conducted in [Hudelot *et al.*, 2008] makes use of structural information for helping the recognition process; fuzzy extensions of the standard RCC properties are used for dealing with medical images, in order to focus on conceptual representations more then on concrete data. In the object recognition phase we use semantic information in a similar fashion, the main differences being that such information is automatically produced (i.e., not provided by a domain expert) and the reasoning is performed via a fuzzy reasoner.

Some results of our research are shown in [Vitucci *et al.*, 2010], in which a common household object is used as an example of the building of a semantic representation.

3 Conclusions and future work

The images and models we are currently using represent synthetic household objects and toys, but we aim to realize a system capable of dealing with real life objects. There are many challenges to face, the most important being an efficient segmentation process and an unambiguous definition of *part*, in order to build reasonable yet effective representations; moreover, the problem of occlusion, i.e., of distinguishing objects which are superimposed on other objects, is critical. To address both problems, we are planning to take advantage of a stereo camera (such as a Kinect) in order to exploit more spatial information and obtain suitable object models after a reconstruction phase. A big challenge will be to effectively implement such a reasoning system on a robot and to perform experiments in a real environment (e.g., a kitchen).

Finally, we are working on how to assess the plausibility and efficacy of a grasping act and include such information in the grasp ontology.

[1] *http://gaia.isti.cnr.it/ straccia/software/fuzzyDL/fuzzyDL.html*

[2] *http://www.image.ece.ntua.gr/ nsimou/FiRE/*

References

[Agarwal *et al.*, 2004] Shivani Agarwal, Aatif Awan, Dan Roth, and IEEE Computer Society. Learning to detect objects in images via a sparse, part-based representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:2004, 2004.

[Biederman, 1987] I. Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2):115–117, 1987.

[Desolneux *et al.*, 2004] A. Desolneux, L. Moisan, and J.-M. Morel. Gestalt theory and computer vision. In A. Carseti, editor, *Seeing, Thinking and Knowing*, pages 71–101. Kluwer, 2004.

[Hudelot *et al.*, 2008] C. Hudelot, J. Atif, and I. Bloch. Fuzzy spatial relation ontology for image interpretation. *Fuzzy Sets and Systems*, 159(15):1929–1951, 2008.

[Huebner *et al.*, 2009] K. Huebner, K. Welke, M. Przybylski, N. Vahrenkamp, T. Asfour, D. Kragic, and R. Dillmann. Grasping known objects with humanoid robots: A box-based approach. In *Advanced Robotics, 2009. ICAR 2009. International Conference on*, pages 1–6, June 2009.

[Keet and Artale, 2008] C. M. Keet and A. Artale. Representing and reasoning over a taxonomy of part-whole relations. *Appl. Ontol.*, 3:91–110, January 2008.

[Lukasiewicz and Straccia, 2008] T. Lukasiewicz and U. Straccia. Managing uncertainty and vagueness in description logics for the semantic web. *Journal of Web Semantics*, 6(4):291–308, 2008.

[Modayil and Kuipers, 2007] J. Modayil and B. Kuipers. Autonomous development of a grounded object ontology by a learning robot. In *Proceedings of the Twenty-Second National Conference on Artificial Intelligence (AAAI-07)*, 2007.

[Rubio *et al.*, 2010] O.J. Rubio, K. Huebner, and D. Kragic. Representations for object grasping and learning from experience. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 1566–1571, October 2010.

[Simou *et al.*, 2007] N. Simou, T. Athanasiadis, V. Tzouvaras, and S. Kollias. Multimedia reasoning with f-$\mathcal{SHIN}$. In *Second Intl. Workshop on Semantic Media Adaptation and Personalization*, pages 44–49, 2007.

[Varzi, 2007] A. C. Varzi. Spatial reasoning and ontology: Parts, wholes, and locations. In M. Aiello, I. E. Pratt-Hartmann, and J. van Benthem, editors, *Handbook of Spatial Logics*, pages 945–1038. Springer-Verlag, 2007.

[Vitucci *et al.*, 2010] N. Vitucci, M. Arrigoni Neri, and G. Gini. Using f-$\mathcal{SHIN}$ to represent objects: An aid to visual grasping. In *UniDL '10*, pages 80–87, August 2010.