

# Object recognition and ontology for manipulation with an assistant robot

Hélène Vorobieva<sup>1</sup>, Mariette Soury<sup>2</sup>, Patrick Hède<sup>1</sup>,  
Christophe Leroux<sup>2</sup>, Philippe Morignot<sup>2</sup>,

<sup>1</sup> CEA, LIST, Vision and Content Engineering Laboratory,  
18 route du Panorama, BP6

F-92265 Fontenay-aux-Roses, France

<sup>2</sup> CEA, LIST, Interactive Robotics Laboratory,  
18 route du Panorama, BP6

F-92265 Fontenay-aux-Roses, France

{helene.vorobieva, , mariette.soury, patrick.hede,  
christophe.leroux, philippe.morignot} @cea.fr

**Abstract.** This article presents a service robotic system for people loosing their autonomy developed at CEA LIST. In the past on SAM robot, we have developed a method for automatic manipulation and object grasping using visual servoing. This method is too stereotyped to correctly grasp objects with complex geometry or to assign particular use to the manipulated object. In this article, we present a new study to adapt the grasping and the usage of an object designed by the user. Our method uses vision object recognition (CBIR) and an ontology for robotic manipulation. This recognition is implemented as a Web Service. It relies on passive vision and does not use a geometric model for grasping. The implementation of this method enables us to automatically search objects in the surrounding areas and to play cognitive and physical stimulation games with the user.

**Keywords:** cognitive robotics, service robotics, manipulation, grasping, handicap, quadriplegia, elderly people, loss of autonomy, object recognition, ontology, Web Service, interoperability, object search, stimulation games.

## 1 Introduction

People loosing their autonomy (disabled, elderly persons) and needing assistance in their everyday life generally resort to caretakers. Nevertheless, some easy and frequent tasks could be done by a service robot in order to give more freedom and autonomy to those people. Among those tasks, there are grasping and manipulation of everyday life objects. Even if a lot of methods exist for grasping, they are generally stereotyped and without a favored and adapted posterior use of the object.

However, we would like for example to be able to ask the robot to bring something to drink, supposing that the drink is in a cup. The robot should then recognize a cup among other objects, associate the cup to the action “give a drink” and grasp it securely (by the side opposite to the handle).

## 1.1 Previous works

Here we shortly describe some main past contributions.

**Grasping knowing the place of the object or with controlled environment.** It can be where the position of each object is *a priori* known (projects RAID [1], DEVAR [2], Master [3]). Environment can also be equipped with intelligent systems such as intelligent tables [4]. Those tactile tables, covered by a sort of artificial skin, allow localization of objects of more than 5 grams. Those methods need a perfectly controlled and equipped environment which can be costly, difficult to generalize and can reduce freedom of robot's actions.

**Grasping with use of 3D geometric model.** A 3D model of object is needed. During the grasping, information from sensors, cameras or lasers is compared with the pre-established model to estimate the position of objects during tracking [5]. This method is used in the project CARE-O-BOT [6]. The development of the 3D model and the information matching during the grasping can be difficult, in particular with the presence of concave and convex regions in the object.

**Grasping without model or object marking.** The user can select an object on a graphic interface by drawing a bounding rectangle as we previously did [7]. Another example is the selection with a laser cursor for the service robot EL-E [8]. Those methods do not allow an adaptation of the grasp strategy or the later use of the object.

## 1.2 Contributions

Here we describe our contribution for object grasping. The method is detailed in section 2. Our method does not need any 3D geometric models even partial. For us, the model of an object is a small group of 2D images. Acquisition of those images does not need technical competences as is required to build 3D models. The objects do not have to be in the vision field of the user as it is the case with a designation of the object by a laser cursor.

Our recognition method uses image indexing and allows estimating the angle or point of view on the object regarding the position of the arm. Once this made, the object grasping strategy is obtained from an ontology, which also contains information on possible usage and type of objects. Unrecognized objects can be always grasped by our previous method [7]. A recognition Web Service using DPWS standard [9] was created for this method to assure interoperability with the services from partners of ITEA MIDAS project<sup>1</sup>. The objective of this European project is the design of a multimodal interface for assistance at home or during driving for people losing their autonomy.

This recognition enabled us to develop an intuitive object selection to ease the use of the interface during object grasping and an object search program. It can also be used for cognitive and physical stimulation games with the user.

The next section presents the robot we are using for the development of the recognition and the interoperability resulting from our implementation. Section 3 deals with the details of object recognition and associated ontology. Our intuitive

---

<sup>1</sup> More details on [http://www.itea2.org/public/project\\_leaflets/MIDAS\\_profile\\_oct-08.pdf](http://www.itea2.org/public/project_leaflets/MIDAS_profile_oct-08.pdf)

object selection is detailed in section 4 and section 5 presents the applications to object search and stimulation games.

## 2 Implementation on the robot and interoperability

This study takes place as part of European project ITEA MIDAS on assistance to people loosing their autonomy. Our team works on home assistance. The robot SAM [7] which we develop (Fig. 1) is meant to stimulate and help people in their everyday life. This means being able to understand the environment and being able to automate as much as possible actions to accomplish. In that respect grasping and manipulation of various objects become essential. Nevertheless the equipment of the robot should stay cheap and easy to use. We use a gripper with stereo camera (for the visual servoing), pressure sensors, optical barrier to detect when the object is in the gripper (Fig. 1).

An intuitive interface allows the user to send the robot to another room, to see its travel with a panoramic camera and then to select an object. This interface was tested during clinical assessment [10] which demonstrated its efficiency, ease of use and the satisfaction of the users towards this type of control.

The object recognition program was developed as a Web Service with a client server structure. When the user wants to select or search for an object, the client sends the current image to the network, the server receives and analyzes it and sends the result back on the network for the client. This method facilitates the interoperability with other software or home automation devices for home or driving assistance: there is no need to insert the whole recognition program, a client is sufficient.

For this Web Service, we resort on the DPWS architecture [9] (Device Profile for Web Service), a communication protocol based on SOAP-XML, which homogenizes the exchanges between the various services connected to the same network. This technology allows the “plug and play” of different services available from the network immediately after connection. DPWS is implemented in C++ and in Java, so programs written in different languages can communicate easily.

This Web Service is one of several with other already developed Web Services to control the robot (mobile platform, arm, user’s interface) [7].

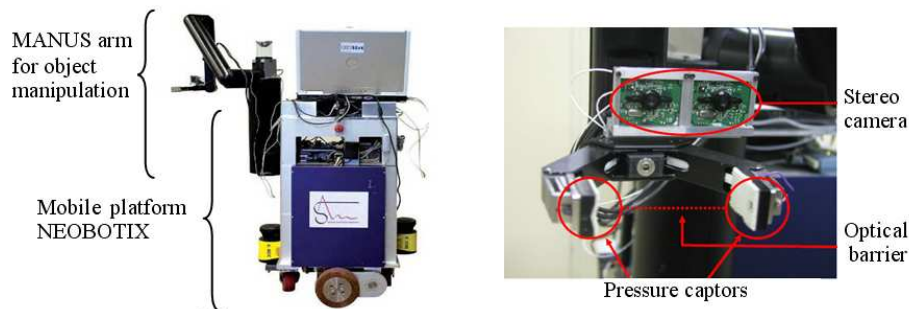


Fig. 1. Service robot SAM (left) and its gripper (right).

### 3 Object Recognition and ontology

#### 3.1 Learning and recognition

To learn an object we need photos corresponding to different points of view on the object (Fig. 2). For each photo the interest points, or keypoints, are extracted using the software ViPR from Evolution Robotics [11] with the SIFT method [12] which relies on difference-of-Gaussian of nearby scales separated by a constant factor  $k$ :

$$DoG(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)} - \frac{1}{2\pi k^2 \sigma^2} e^{-(x^2+y^2)/(2k^2\sigma^2)}. \quad (1)$$

A point of view on the object is described by a vector of coordinates of the keypoints and their texture. The name of those photos is chosen to ease use in an ontology.

This database is easy to create and does not need specific competences. Indeed, the images (2D photos) can be done putting the object on a turntable like in [13] and taking photos of object's views with nearby camera. With a motorized turntable and automatic photo capture, new objects can be easily learned to complete the database.

During the recognition, ViPR extracts the keypoints from the image and compares their feature vectors with those of the database to find potential object matches [11]. Several objects can be identified in one image, including partially occluded objects, if there are at least 4 keypoints (Fig. 3). This recognition is robust to variations of luminosity and can be used in non uniform lightened places.

For this recognition, we first need to load the database (only once). Using our Web Service it takes 15,5s for a database of 72 images on a PC Intel Core 2 CPU 2,66 GHz 3,50 Go RAM. Next, the Web Service can recognize the objects (average 450ms).



**Fig. 2.** Different points of view of some objects from the database and extracted keypoints. Plastic bottle (*left*), box of chocolate milk mix (*center*), cup (*right*).



**Fig. 3.** Keypoints corresponding to different recognized objects: box of chocolate milk mix (full white circles), box of sweet (empty circles), pepper pot (full black circles).


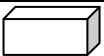

### 3.2 Ontology

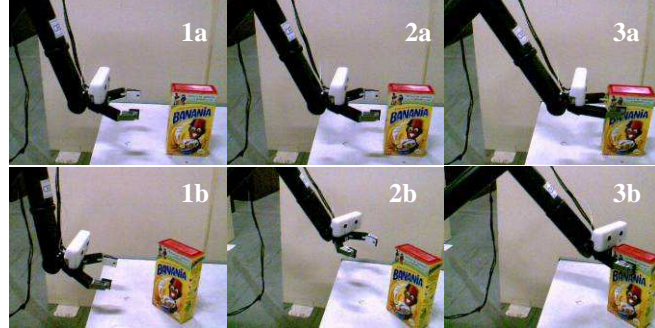
Ontology in computer science is a concept used for knowledge representation i.e. objects and concepts of a domain and the relations between them. It allows a level of abstraction of data models with a more semantic representation [14].

To find which grasping or object manipulation strategy to use, we create an ontology for robotics manipulation with XMLSpy. The ontology contains grasping strategies suited to each image or group of images from the database, according to the point of view on the object and its geometric structure (Table 1). It includes particularly the moves to make to position the gripper in an adapted place to grasp the object according to the morphology of the object and of the gripper (Fig 4). When the gripper has reached this place, it has only to move forward and do a blind grasp. All those moves are done after a visual servoing as in [7] in order to be always at the same distance from the object before the beginning of the motion from the strategy.

The ontology can also contain information about pressure to apply on the object during the grasping, about use of the objects: for example, the action “drink” can be associated to containers (cup, can), the concept “breakfast” – to coffee and cereals boxes, the place “bathroom” – to toothpaste (probable place where this object is). This information can be used for an oriented research of an object.

**Table 1.** Examples of grasp strategies in our ontology.

Name of the strategy	Object's geometry	Possible objects	Angle, point of view
RevolutionSymetry		Can, bottle, glass	indifferent
RectangularCuboid000		Box of pills, box of cereals	0° or 180°
RectangularCuboid045			45° or 225°
Cup000		Cup	0°
Cup045			45°
Cup090			90°



**Fig. 4.** Examples of grasp strategy: (a) RectangularCuboid090: box seen with an angle of  $90^\circ$ , the gripper moves straight forward ; (b) RectangularCuboid045: box seen with an angle of  $45^\circ$ , moving of the arm in left direction, modification of orientation of the gripper before moving forward.

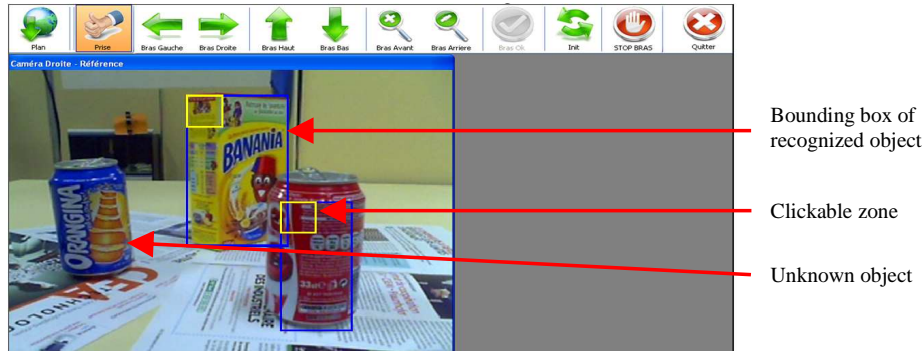
## 4 Intuitive object selection

Before grasping an object, we use a visual servoing [7] to place the arm in front of the object. As explained before, this action is essential for the correct grasping. For this servoing we need to select the object in a bounding box. Previously, this selection was done in 2 mouse clicks which defined the opposite corners of a bounding box containing the object [7] [10]. This method requires only 2 clicks but for disabled persons every action can take a lot of time and effort: inaccurate clicks because of cognitive difficulties, physical difficulties which require specific equipment instead of mouse. So, when the object is known in the database, we want to reduce even more the number of actions for the object selection.

When an object is recognized, we know the position of the recognized points of interest  $P_{x,y}$  and define the bounding box thanks to those points. We noticed that those points reach rarely the edges of the object so we decided to enlarge the bounding box extremities with an empirically defined constant 'e':

- left top corner of the box  $[\max(x_{\min}-e,0),\max(y_{\min}-e,0)]$
- right bottom corner of the box  $[\min(x_{\max}+e,width_{image}),\min(y_{\max}+e,height_{image})]$

When the user wants to choose an object in the scene, all the recognized objects are shown with their bounding box so the user only has to click in the desired bounding box. To prevent object box superposition, we decided to reduce the clickable zone for this selection (Fig. 5). If the object that the user wants to select is not recognized, he can define a bounding box by 2 clicks as it was done before. So when the user clicks on a clickable zone, the recognized object is selected, when the user clicks otherwise, this click defines one of the corners of a bounding box.



**Fig. 5.** Interface during object selection. Recognized objects are in bounding boxes (*blue*) and can be selected with a click in the clickable zone (*yellow*). Unrecognized object can be selected by two clicks.

## 5 Applications to object search and user's stimulation

Since selection of known objects in a bounding box is now automatic, we have implemented an object search in the environment. The user asks the robot to find an object (in a list of known objects) and the robot travels in the environment until he has found the object or has searched in all possible places. If the object is found, it can be automatically brought to the user. For example, we have assigned to each possible station (like tables) different positions of the arm to glance over all the surface of the station. We have tested our search program for one of those stations (120x100 cm). When the objects presents a lot of points of interest (big and textured objects), the research has a good success rate. This rate decreases with the number of detected points of interest (Table 2).

Assistance robotics can also be preventive or stimulating regarding cognitive or physical state of users. So with the object recognition, we can create stimulating games (for children or people with Alzheimer's disease).

For example the robot asks the user to show him one by one a set of known objects and places the arm every time in a different position. The stimulation is cognitive because the user has to find the right object and physical because the user has to reach the camera on the gripper in order to show the object and validate this task.

**Table 22.** Success search rate for different types of objects. Occultation decreases the number of interest points available for the recognition during object research.

Type of object		Big, textured	Big, little textured	Small, textured	Small, little textured
Success search	No occultation	86%	64%	68%	42%
	Occultation	71%	42%	29%	25%

## 6 Conclusion and future work

This article presented a new study on assistance for object grasping and manipulation. A 2D object recognition allows an adaptation (movement, pressure of grasping) of object grasping, but still allows grasping unknown objects. This vision method does not need geometric model of the object and is robust. Object learning is easy and can be done without robotics knowledge. All types of objects can be recognized if they have enough texture. Thanks to this recognition, we can grasp objects which could not be grasped before (such as box which width is bigger than gripper and seen full-frontal). The object selection by the user is eased and the robot can autonomously search objects. Cognitive and physical stimulation games can be implemented. The recognition allows the amelioration of the robot's environment knowledge and represents one more step toward intelligent and autonomous object manipulation.

We are currently working on the elaboration of new assistance scenarios and stimulation games. The ontology will be completed with an association of probable places for different objects to make the object search faster. Plan generation will soon give the possibility to user's assistants to define by themselves individualized scenarios. Clinical assessment of this method is planned in the ITEA MIDAS project.

## References

1. Dallaway, J., Robin, S.: Raid -a vocational robotic workstation. In: IEEE ICORR. RU (1992)
2. Van der Loos, H.: Va/stanford rehabilitation robotics research and development program: Lessons learned in the application of robotics technology to the field of rehabilitation. In: IEEE Trans. on Neural Systems and Rehabilitation Engineering, pp. 46--55 (1995)
3. Busnel, M., et al.: The robotized workstation "MASTER" for users with tetraplegia: Description and evaluation. *Journ. of Rehabilitation Reseach &Development*, 36(3) (1999)
4. Volosyak, I., Ivlev, O., Gräser, A.: Rehabilitation Robot FRIEND II The General Concept and Current Implementation. In: Proc. of the 2005 IEEE ICORR. Chicago, IL, USA (2005).
5. Bourgeois, S., Naudet-Collette, S., Dhome, M.: Recalage d'un modèle CAO à partir de descripteurs locaux de contours. In: RFIA. Tours, France, (2006)
6. Graf, B., Hans, M., Schraft, R.: Care-o-bot II development of a next generation robotic home assistant. *Auton. Robots*, 16(2), pp. 193--205 (2004)
7. Remazeilles, A., Leroux, C., Chalubert, G.: SAM: a robotic butler for handicapped people. In: IEEE RO-MAN. Munich, Germany (2008)
8. Jain, A., Kemp, C. C.: EL-E: An Assistive Mobile Manipulator that Autonomously Fetches Objects from Flat Surfaces. In: *Autonomous Robots, Special Issue* (2009)
9. Jammes, F., Mensch, A., Smit, H.: Service-oriented device communications using the devices profile for web services. In: AINA Work. pp. 947—955. Washington, USA (2007)
10. Leroux, C., et al.: Robot grasping of unknown objects, description and validation of the function with quadriplegic people. In: IEEE ICORR. Noordwijk, Pays-Bas (2007)
11. Karlsson, N., et al.: Core technologies for service robotics. In: IROS (2004)
12. Lowe, D. G.: Object recognition from local scaleinvariant features. In: *International Conference on Computer Vision*. pp. 1150—1157. Corfu, Greece (1999)
13. Nene, S. A., Nayar, S. K., Murase, H.: Columbia Object Image Library: COIL-20. Technical Report CUCS-005-96. Columbia University (1996)
14. Gruber, T.: *Ontology. Encyclopedia of Database Systems*, Ling Liu and M. Tamer Özsu (Eds.), Springer- Verlag (2009)