Proceedings of the 2006 IEEE
International Conference on Robotics and Biomimetics
December 17 - 20, 2006, Kunming, China

3D Grasp Synthesis Based on Object Exploration*

Eris Chinellato, Gabriel Recatalá and Angel P. del Pobil
Robotic Intelligence Lab
Dept of Computer Science and Engineering
Jaume I University
12071 Castellón, Spain
grecata,eris,pobil@icc.uji.es

Youcef Mezouar and Philippe Martinet
LASMEA
Blaise Pascal University
63177 Aubière, France
mezouar,martinet@lasmea.univ-bpclermont.fr

***Abstract*— Many approaches to robotic grasping have focused on a specific aspect of the problem only, without considering its integrability with other related procedures in order to build a more complex task. The model for grasp synthesis presented in this paper, inspired on human neurophysiology, is built upon an architecture that allows its scalability and its integration within more complex tasks. The grasp synthesis is designed as integrated with the extraction of a 3D object description, so that the object visual analysis is driven by the needs of the grasp synthesis: visual reconstruction is performed incrementally and selectively on the regions of the object that are considered more interesting for grasping. Our approach, inspired by the efficiency of our visual cortex, allows for an easy integration of additional modules and different grasp synthesis criteria.**

***Index Terms*— biologically-inspired robots, active perception, biomimicking robots/systems, grasping/dexterous manipulation, robot vision**

I. INTRODUCTION

The ability to manipulate every kind of objects in a dexterous way is one of the most distinctive abilities in humans, and also one of the fundamental skills pursued by robotic researchers. Nevertheless, despite the amount of research and technological efforts, there are still important differences between humans and robots that influence the way robotic grasping applications can be defined. In the first place, human hands are characterized for having five soft fingers with high dexterity and compliance, whereas robotic hands feature a lower level of dexterity and less elaborated contact surfaces. In addition, the human brain has a degree of parallelism much higher than any ordinary current computer. Finally, the action of manipulating objects in humans involves the control of a number of elements –hand, arm, eyes, head– that have, globally, many more degrees of freedom than current robotic setups can manage. Therefore, neuroscience models of the flow and processing of information in the brain of humans and other primates cannot be directly applied to a real robotic system, but have to be adapted, or tailored, to it.

The problem of selecting the way to grasp an object with a robotic hand has been widely analyzed in the literature. In the case of considering a 3D object description, the grasp search has been performed in many works on a model of the object. Although many solutions exist for the 3D reconstruction of objects and scenes from visual data, the integration of this reconstruction with some task oriented-processing, such as the grasp search, has not been fully developed yet. In fact, many works regarding these problems have not considered the *integrability* of their solutions with other related procedures in order to build a more complex task.

One of the contributions of this paper is the adaptation of part of a model of information processing for vision-based grasping in the human brain [1] to a robotic system. In addition, an architecture is proposed for the development of the above model, following behavior-based guidelines. This architecture supports the nesting and concatenation of processing modules in a structured way.

The grasp planning we propose is formulated through the integration of object visual analysis and grasp search procedures. In particular, the grasp-synthesis method uses a multi-resolution representation of the object, in which the 3D visual analysis procedure depends on criteria for selecting, and reconstructing more thoroughly, the object features that appear more relevant for grasping purposes. Therefore, the incremental, selective 3D object reconstruction is obtained through action-oriented object exploration.

In the next sections, after presenting related research and neuroscientific background, we will introduce our architecture, describe its application to the vision-based grasping model, and show some experimental results.

II. RELATED WORK

The study of the areas of the human brain involved in the different stages of a manipulation task, and of the flow of information through these areas, is an important and ever developing field in neuroscience [2]. The exact nature of this information and how it is elaborated is hot-topic in neuroscience; new light is being shed especially regarding the role of the associative cortex which makes use of visual data processing for action planning. The modeling of the link between visual and associative visuomotor cortex is the object of this work. One of the most important related findings is the duality between vision for action and vision for perception. In fact, the human visual system is made out by two main information streams, a *dorsal pathway*, more oriented toward

*This work has been partially supported by Generalitat Valenciana (GV05/137, CTIDIA/2002/195), by Fundació Caixa-Castelló (P1-1A2003-10 and 05I006.30) and by the Ministerio de Educación y Ciencia (DPI2001-3801, DPI2004-01920, FPI grant BES-2002-2565).

1-4244-0571-8/06/$20.00 ©2006 IEEE

action-based vision, and a *ventral pathway*, more suitable to categorization and recognition tasks [3]. Although the dorsal stream is the one more critical for the planning and execution of grasping actions, to the extent that some of its areas are especially dedicated to this task, only the interaction between and within both streams can lead to the complex and reliable human grasping skills [4].

Intents to emulate the computation of the visual cortex have been carried on, and models of grasping are also available, as for example the FARS model [5], which focuses especially on the action-execution step. Nevertheless, no robotic applications have been yet developed following this path, and the integration between the two visual pathways is nearly unexplored. More details on the neuroscience findings which inspired this work are found in [1], including hypothesis on how the two streams of visual processing in our cortex coordinate and interact.

In the engineering literature, the grasp stability has often been evaluated in terms of *force* and *form closure* conditions, which ensure stability assuming point contacts with friction [6]. For simplicity, some works have considered an object description that was restricted to lie on a plane [7]. Works considering a 3D object description have normally used a model of the object, obtained or defined in an off-line stage [8], [9]. Alternatively, some authors use heuristics to reduce the number of candidates during the grasp synthesis and obtain a good grasp in short time [10]; others approximate an object model with a set of shape primitives –such as cylinders, boxes, or cones– and use rules, based on those primitives, to generate grasp pre-shapes and starting positions [11]. Quality measures for the analysis grasp analysis have been described in several works [12]. Among vision-based works, in many cases there is a 2D grasp synthesis in one image, followed of a 3D reconstruction and/or validation [13]. Nevertheless, in spite of the number of grasp synthesis methods available, most works on robot-to-object positioning for grasping have used features other than the grasp points.

However, the 3D reconstruction of the model of an object based of visual information is a relatively complex task. Although the use of feature correspondences between several images has been common in many works, such correspondences are not always available, for instance, when the objects have smooth surfaces, so procedures not requiring them have also been developed [14]. In general, the 3D reconstruction produces either a surface-based representation of the object [15] or a volumetric representation [16].

Finally, when trying to develop systems which model or imitate natural skills, two approaches are usually cited as contrary options: bottom-up and top-down. Modern robotics favors the former, which is more plausible from a physiological point of view [17]. Behavior-based systems have proved their effectiveness in dealing with dynamic and complex environments, and have been widely used many fields of robotics. Although this paradigm has been more widely used in other areas of robotics, it is being increasingly used in the development of manipulation tasks [18].

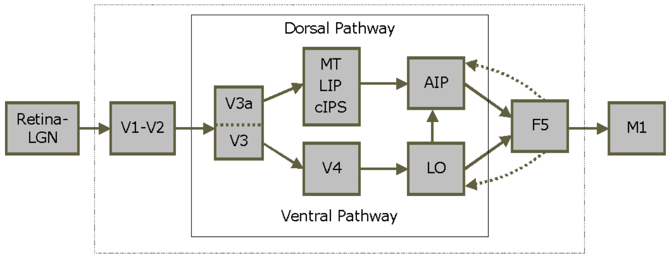

Fig. 1. Block diagram of a grasping system based on human physiology.

III. THE VISION-BASED GRASPING MODEL

In this work, we have considered part a model of the processing of visual information in human grasping, based on recent studies from neuroscience. The part of the model dedicated to the on-line grasp synthesis based on visual information is developed in further sections in order to tailor it to a robotic system.

In the framework outlined in figure 1, the visual input is processed in two parallel ways, one (ventral) more concerned with perceptive information about the object nature, the other (dorsal) oriented to spatial analysis. The products of the visual analysis are, from the dorsal elaboration, precise information about position and geometry of the object, and, from the ventral elaboration, data about expected weight, friction, and previously experienced grasping actions on such object. Blocks in this model have been labeled after the name of the brain regions their behavior is associated to. More details can be found in [19].

From a practical point of view, the main novelty of such a framework is the parallel computation performed on the visual input, so that different aspects are taken separately into account, and the elaboration is carried on according to the purpose of each pathway. At the moment of selecting and planning the most appropriate action, each pathway contributes with the results of its elaboration, and the on-line data gathered and processed by the dorsal stream can be complemented and enriched by the previously stored knowledge recovered by the ventral stream. In this work we start from the hypothesis that the object to grasp is unknown, so that the on-line visual analysis is prevailing. Thus, we dedicate most attention to the dorsal stream, which is likely to identify possibly graspable features on target objects and through visual exploration refine and complement the available data, until it is considered reliable enough to proceed to the synthesis of an executable grasp.

The diagram of figure 1 is very simplified, but it is not our purpose here to develop a full, distributed model of the visual cortex. Instead, we only take into account the main task of each area and its most probable connections to nearby areas, in order to reconstruct a simplified, sequential information flow related to grasping.

IV. THE FILTER-BASED ARCHITECTURE (FBA)

In this section, an architecture is proposed that supports the development requirements of the model described in the

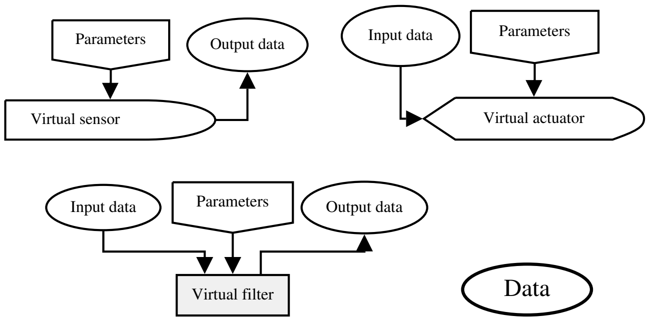

Fig. 2. Basic components of the FBA architecture.

previous section. This architecture, intended for its application on artificial devices, has some intrinsic limitations with respect to the biological model, which influences aspects such as the degree of parallelism and the flow of information. Our architecture uses the following basic types of components, shown in figure 2:

- *Virtual sensors.* Components that provide data acquired from real sensors installed in the system (cameras, infrared cells, etc.). They correspond to the primary sensory areas of the human cortex (rather than to the sensory organs).
- *Virtual actuators.* These components receive commands or data to be sent to physical/real actuators installed in the system (robot arm, gripper, etc.). They model the primary motor areas.
- *Virtual filters.* These components process the data they receive from virtual sensors and/or other filters, and produce some results, which are provided to other filters or to virtual actuators. They handle operations such as feature extraction or a control law. Each filter can be seen as a specific cortical area. In the brain, they are disposed more as a continuous than as in a block diagram, but technological constraints oblige to build a simpler model, in which areas are separated from each others and connected through a clear input-output flow.
- *Data sets.* They constitute groups of data that are produced and processed by the above modules. They represent an over-simplification of the information flow connecting brain areas.

Virtual sensors, actuators and filters have interfaces, through which they are interconnected. As shown in figure 2, three types of interfaces are considered:

- *Input interface.* It indicates the set of data that a given component requires as input.
- *Output interface.* Specification of the set of data that a given component provides as output.
- *Parameter interface.* Set of parameters that can be used to configure a component.

A *task* will be the set of all connected virtual sensors, filters and virtual actuators that are simultaneously active within a system. The data sets constitute an internal, non-centralized memory spread along the chain of components.

V. The grasp-synthesis model

According to the FBA architecture outlined in the previous section, the model presented in figure 1 can be developed as in figure 3 using the blocks of figure 2. This paper focuses on one specific task, the on-line grasp visual synthesis performed by the dorsal stream in conjunction with the visual areas, through active visual exploration of target objects [4].

A. The grasp-synthesis task

The experimental setup considered in this work consists of a robot arm equipped with a camera mounted in an eye-in-hand configuration. A three-fingered hand has been considered for the synthesis and execution of grasps. The objects should be of a graspable size and shape for this hand, and with a convenient texture, so that its contour can be extracted by the vision system in each acquired image.

The proposed exploration is active, guided by the need of searching or computing specific data that are required for the grasp synthesis. This strategy constitutes a general framework within which different grasp synthesis and analysis criteria could also be tested.

The robot arm is intended to perform a movement around the selected object during which images are acquired. However, other configurations could also be valid, as long as the exploratory movement can be performed.

Figure 4 provides a general description of this task, which is composed of the following steps:

- *Visual analysis*. Initial stage in charge of gathering the visual data necessary to start the grasp synthesis and the object exploration.
- *Grasp synthesis*. This is a deliberative stage, in which the actual grasp synthesis is performed, based on the initial data set and, mainly, the data collected during the exploratory movements. If additional information is required to perform the grasp synthesis, a plan is made for a new exploratory movement.
- *Exploration movement*. In this stage, the system performs some planned exploratory movement in order to extract new information about the object to improve the data available to the grasp synthesis process.

The stages are iteratively executed until the grasp synthesis algorithm is able to select a grasp or it decides to cancel the grasp search because of a failure.

In the following sections, we focus on the vision analysis issues associated to the grasp synthesis, rather than on the control and planning of the exploratory movement.

B. Action-oriented visual analysis

This stage roughly corresponds to the tasks performed by the advanced visual areas V3-V3a and the first associative parietal areas, as cIPS. The goal is to extract the visual information required to identify and evaluate object features useful for grasping and start planning exploratory movements that may help in the grasp synthesis. This stage is left when either the initial data extraction is finished (and the dorsal

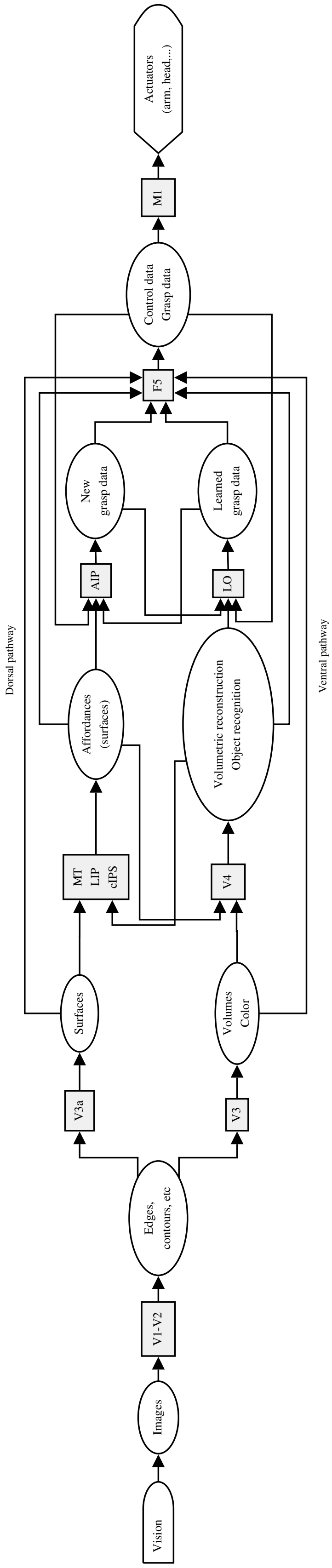

Fig. 3. Development of the grasping model from figure 1 using the FBA architecture.

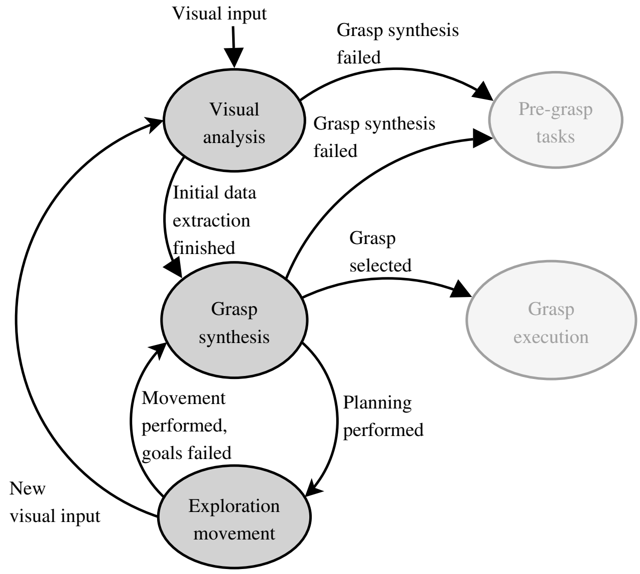

Fig. 4. Automata-based description of the grasp-synthesis task.

stream comes into action) or it has not been possible to perform it (for example for a sudden occlusion).

In our development, the initial information consists of a rough 3D reconstruction of the object, with the purpose of identifying the possible grasp zones. In order to obtain a fast reconstruction at this stage, only a reduced number of views of the object is used. The visual analysis is performed using the contours of selected objects, extracted by the modules in charge of basic image processing (corresponding to the primary visual cortex).

The analysis is based on the use of an octree structure, which allows to control its degree of detail. Algorithm 1 describes the general procedure for carving the octree. This algorithm is applied to each input contour-based object description. The level of detail of this reconstruction is controlled by limiting the maximum number of levels in the octree and the minimum size of the octree nodes. The decision on whether to carve an octree node or not is determined by algorithm 2; for this stage, the threshold is set to zero.

Algorithm 1 Octree-based reconstruction

```
Initialize root voxel of the octree
for each level in the octree do
    if addition of a new octree level is allowed then
        for each node at this level do
            if carving of this node is required then
                Create child nodes
            end if
        end for
    end if
end for
```

Algorithm 1 requires some estimation of the volumetric size of the object, in order to initialize the root voxel of the octree. This information can be previously known or estimated using some fast point-reconstruction method.

Algorithm 2 Get node carving requirement

Project node points on the image
voxel occupancy $\leftarrow$ % projected points inside obj. contour
carving required $\leftarrow$ false
if voxel occupancy = 0 **then**
  if overlapping between object and node projection **then**
    carving required $\leftarrow$ true
  end if
else if voxel occupancy $\leq$ threshold **then**
  carving required $\leftarrow$ true
end if

C. Grasp synthesis

This is the stage in which the generation of the grasp to execute is actually performed, using data coming from the visual analysis stage during the object exploration process. In this deliberative stage, the object data are analyzed in order to select positions for placing the fingers of the robot arm. If the available information is considered enough for the synthesis of a reliable grip, the control passes to the premotor areas in charge of organizing the target movements; otherwise, a new exploratory movement is performed and new visual data gathered and added to those already available.

In this work, the selection of grasps has been performed using the grasping simulator "GraspIt!" [11]. The octree-based object model produced by the initial and exploratory stages is provided as input to this simulator. This software includes a grasp planner consisting of two components, one for generating a set of starting grasp locations on the object model and another one for testing and evaluating them.

D. Exploration movement

The activity in this stage is oriented to the extraction of information about the object during the execution of a planned exploration movement. The information extracted from each acquired image is used to incrementally update the object description produced by the visual analysis stage.

The overall behavior can be briefly described as follows:

- *Visual assessment of available feature data.* The grasp generation stage, in its intent of generating possible grips, assesses the quality of the available data through the use of visual criteria [20]; based on these data, it parameterizes the further exploration of areas that are considered interesting for grasping but not reliably covered by the visual analysis.
- *Movement control.* The movement is performed in order to optimize the visual knowledge referred to object parts more relevant for action. In our cortex, such a movement could only be obtained by the integration of parietal and vestibular information, as it requires control of the head and various different transformations of the reference frame. In our case, the visual system is joined and controlled together with the hand, so that we only control a hypothetical premotor cortex responsible for the joint movement of sensor and effector.
- *Enrichment of visual knowledge.* Partial reconstruction, oriented to those of portions of the object surface that have been assessed more appropriate for grasping, is realized and added to the information already available in the dorsal areas (cIPS-AIP), which can repeat the synthesis-exploration cycle until a reliable grasp configuration is found.

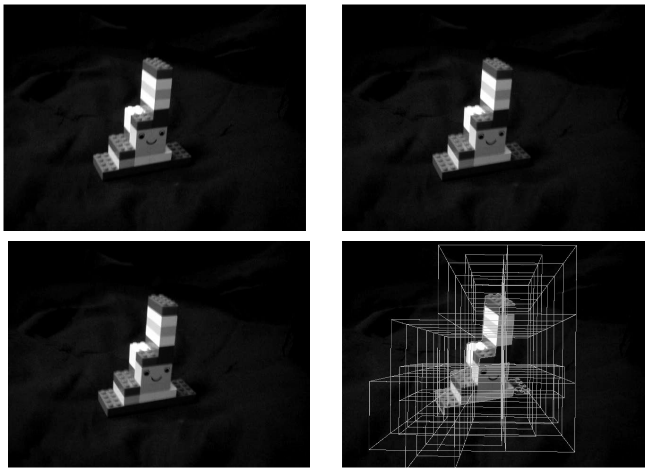

Fig. 5. Initial data extraction. Sequence of object views, with rough object reconstruction in the fourth view.

Algorithms 1 and 2 are used for the refinement of the object reconstruction, according to the evaluation provided by the visual criteria.

VI. Results

The images acquired during the exploratory movement of the robot arm are processed according to the procedures described in section V. Figure 5 shows the results of the *visual analysis* stage based on the initial views of a selected object. Using an octree structure, a rough reconstruction is performed that will be used afterward for the grasp-based object exploration. The generated octree has a maximum of four levels of depth. This reconstruction has been performed with only three images. The size of the root voxel of the octree has been estimated from a set of reconstructed points.

A refinement of the above octree is performed in the *exploration movement* stage during a movement of the vision system around the object. As indicated in section V, a grasp-specific, partial visual analysis is performed. This specific refinement favors the reconstruction of the parts of the object that are likely to be smoother and, therefore, more suitable for placing the gripper fingers for grasping. Figure 6 shows the result of applying this refinement to the rough reconstruction given in figure 5. A generic refinement is performed first in order to better define the shape of the object. Then, more specific criteria intervene so that the reconstruction focuses on those areas of the object that are more interesting for grasping.

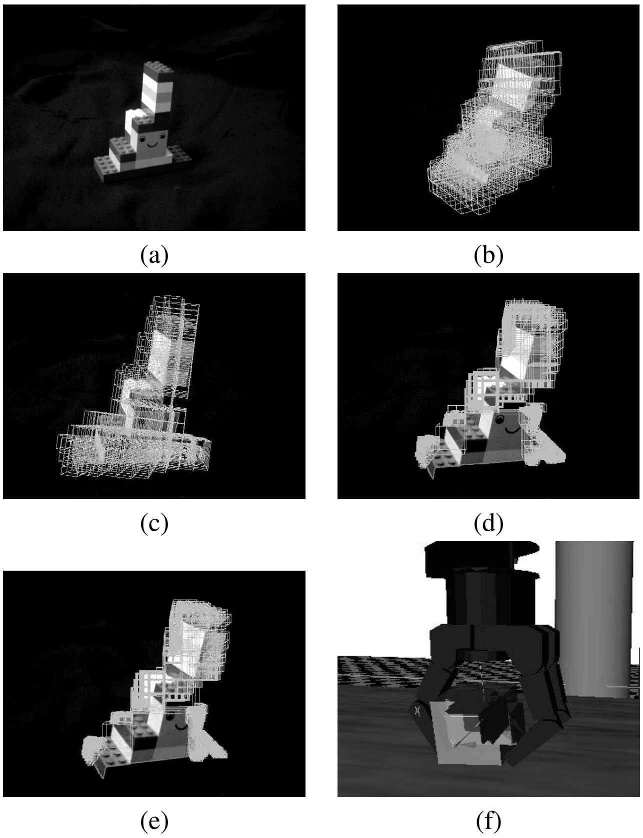

Fig. 6. Exploration movement for grasp synthesis. View after the initial object reconstruction (a), which triggers a generic reconstruction (b)-(c), followed by a grasp-specific reconstruction (d)-(e) and grasp selection (f).

The regions of the object that, based on the above criteria, are considered less suitable are ignored or reconstructed with less detail.

The grasp is selected on the reconstructed part of the object using the simulator "GraspIt!". Figure 6-f provides an example of grasp automatically produced by this software on the given object. A simplified multi-resolution reconstruction of the object has been used in this case.

VII. Conclusion

The first goal of this paper is that of providing a framework for the development of robotic applications on the synthesis and execution of grasps, taking inspiration from the computations performed by our brain when performing this kind of actions. The proposed grasp synthesis can be extended within this framework to obtain a more detailed development of brain models. In addition, since it is based on a grasp-driven multi-resolution reconstruction of the object, future work will have to consider the cases in which it is difficult to extract relevant information from low-resolution reconstructions in the earlier stages of the grasp search.

Finally, although the models described in this paper have been considered from a general point of view, they are oriented to be used in a relatively autonomous system, which would have to handle on its own the execution of specified tasks of a certain degree of complexity. Such a system would be able to act as an assistant, requiring only high-level indications from a user.

References

[1] E. Chinellato and A. del Pobil, "Vision and grasping: Humans vs. robots," in *Lecture Notes in Computer Science*, J. Mira and J. Álvarez, Eds. Springer Verlag, 2005, no. 3561, pp. 368–377.

[2] U. Castiello, "The neuroscience of grasping." *Nat Rev Neurosci*, vol. 6, no. 9, pp. 726–736, Sept. 2005.

[3] M. A. Goodale and A. D. Milner, "Separate visual pathways for perception and action." *Trends Neurosci*, vol. 15, no. 1, pp. 20–25, Jan. 1992.

[4] J. Bullier, "Integrated model of visual processing." *Brain Res Brain Res Rev*, vol. 36, no. 2-3, pp. 96–107, Oct. 2001.

[5] A. Fagg and M. Arbib, "Modeling parietal-premotor interactions in primate control of grasping," *Neural Networks*, vol. 11, no. 7–8, pp. 1277–1303, Oct. 1998.

[6] A. Bicchi and V. Kumar, "Robotic grasping and contact: A review," in *Proc. IEEE Intl. Conf. on Robotics and Automation*, Apr. 2000, pp. 348–353.

[7] J. Ponce, D. Stam, and B. Faverjon, "On computing force-closure grasps of curved two dimensional objects," *The Intl. J. of Robotics Research*, vol. 12, no. 3, pp. 263–273, June 1993.

[8] E. Lopez-Damian, D. Sidobore, and R. Alami, "Grasp planning for non-convex objects," in *Proc. of the 36th International Symposium on Robotics*, Tokyo, Japan, Nov. 2005.

[9] S. Arimoto, M. Yoshida, and J.-H. Bae, "Stable "blind grasping" of a 3-D object under non-holonomic constraints," in *Proc. IEEE Intl. Conf. on Robotics and Automation*, Orlando (Florida), USA, May 2006, pp. 2124–2130.

[10] C. Borst, M. Fischer, and G. Hirzinger, "A fast and robust grasp planner for arbitrary 3D objects," in *Proc. IEEE Intl. Conf. on Robotics and Automation*, Detroit (Michigan), USA, May 1999, pp. 1890–1896.

[11] A. Miller, S. Knoop, H. Christensen, and P. Allen, "Automatic grasp planning using shape primitives," in *Proc. IEEE Intl. Conf. on Robotics and Automation*, Taipei, Taiwan, Sept. 2003, pp. 1824–1829.

[12] C. Borst, M. Fischer, and G. Hirzinger, "Grasp planning: How to choose a suitable task space," in *Proc. IEEE Intl. Conf. on Robotics and Automation*, New Orleans (Louisiana), USA, Apr. 2004, pp. 319–325.

[13] A. Hauck, J. Rüttinger, M. Sorg, and G. Färber, "Visual determination of 3D grasping points on unknown objects with a binocular camera system," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, Kyongju, Korea, 1999, pp. 272–278.

[14] G. Zeng, S. Paris, M. Lhuillier, and L. Quan, "Study of volumetric methods for face reconstruction," in *IEEE Intelligent Automation Conference*, 2003.

[15] F. Chaumette, S. Boukir, P. Bouthemy, and D. Juvin, "Structure from controlled motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 5, pp. 492–504, May 1996.

[16] P. Mendonça, K.-Y. Wong, and R. Cipolla, "Epipolar geometry from profiles under circular motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 604–616, June 2001.

[17] R. Brooks, *Cambrian intelligence: The early history of the new AI.* Cambridge (Massachusetts), USA: MIT Press, 1999.

[18] R. Zöllner, M. Pardowitz, S. Knoop, and R. Dillmann, "Towards cognitive robots: Building hierarchical task representations of manipulators from human demostration," in *Proc. IEEE Intl. Conf. on Robotics and Automation*, Barcelona, Spain, Apr. 2005, pp. 1547–1552.

[19] G. Recatalá, E. Chinellato, A. del Pobil, Y. Mezouar, and P. Martinet, "3D grasp synthesis based on a visual cortex model," in *Proc. of the First IEEE / RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics*, Pisa, Italy, Feb. 2006.

[20] E. Chinellato, A. Morales, R. Fisher, and A. del Pobil, "Visual features for characterizing robot grasp quality," *IEEE Transactions on Systems, Man and Cybernetics, Part C*, vol. 35, no. 1, pp. 30–41, Feb. 2005.