
Biologically-Inspired 3D Grasp Synthesis Based on Visual Exploration

Gabriel Recatalá · Eris Chinellato · Ángel P. del Pobil · Youcef Mezouar · Philippe Martinet

Abstract Object grasping is a typical human ability which is widely studied from both a biological and an engineering point of view. This paper presents an approach to grasp synthesis inspired by the human neurophysiology of action-oriented vision. Our grasp synthesis method is built upon an architecture which, taking into account the differences between robotic and biological systems, proposes an adaptation of brain models to the peculiarities of robotic setups. The architecture modularity allows for scalability and integration of complex robotic tasks. The grasp synthesis is designed as integrated with the extraction of a 3D object description, so that the object visual analysis is actively 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.

Keywords biologically-inspired robots · models of human manipulation · robot vision · grasping/dexterous manipulation · active perception

1 Introduction

The ability to manipulate every kind of objects in a dexterous way is one of the most distinctive in humans, and also one of the fundamental skills pursued by robotic researchers.

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

G. Recatalá (✉) · E. Chinellato · A.P. del Pobil
Robotic Intelligence Lab, Dept. of Computer Science and Engineering,
Jaume I University, 12071 Castellón, Spain
E-mail: grecata,eris,pobil@icc.uji.es

Y. Mezouar · P. Martinet
LASMEA, Blaise Pascal University, 63177 Aubière, France
E-mail: mezouar,martinet@lasmea.univ-bpclermont.fr

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. First, 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, 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.

We know from neuroscience that the human visual system is made out by two main information streams, a *dorsal pathway*, more oriented toward action-based vision, and a *ventral pathway*, more suitable to categorization and recognition tasks [16]. Only a proper interaction between and within both streams can lead to the complex and reliable human grasping skills in complex environments. Nevertheless, the dorsal stream is more critical for the planning and execution of grasping actions, to the extent that some of its areas are especially dedicated to this task. In this work, we present the adaptation of part of a model of information processing for vision-based grasping in the human brain [10] to a robotic system. An architecture is proposed for the development of such model following behavior-based guidelines. This architecture supports the nesting and concatenation of processing modules in a structured way. The implementation of the section of the model more specifically dedicated to grasp synthesis is presented in this paper.

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 thus analyzing more thoroughly, the object features that appear more relevant for grasping purposes. Therefore, the incremental, selective analysis of relevant object features is obtained through action-oriented visual exploration.

This paper is organized as follows. Section 2 provides a brief overall description of works related to the approach considered in this paper. Section 3 includes the definition of an architecture for supporting the adaptation of neuroscience models to robotic setups, and the description of our model for visual grasp search and execution. The implementation of a grasp synthesis task based on visual search taken from such model is given in Section 4, and the obtained results are provided in Section 5.

2 Related work

This section provides a brief insight on the advances in different areas related to this paper. First, a description is given regarding recent studies in the field of neuroscience about the mechanisms of grasping in humans and other primates. Next, works from an engineering point of view are considered, focusing on those related to grasp generation and the extraction of action-related object descriptions. Some references are also given regarding the use of behavior-based architectures for the development of robotic applications.

2.1 Neuroscience

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 rapidly developing field in neuroscience [8]. The exact nature of this information and how it is elaborated is hot-topic and new light is being shed especially regarding the role of the associative cortex which makes use of visual and proprioceptual data for action planning. The link between visual and associative visuomotor cortex is the target of this work.

Visual data in primates flows from the retina to the thalamus, and then mainly (but not exclusively) to the primary visual cortex (V1) in the occipital lobe. There are two main visual pathways going from the basic visual areas V1 and V2 to different cortical association areas: the posterior parietal cortex (dorsal direction) and the inferior temporal cortex (ventral direction) (Figure 1).

Along the dorsal pathway, the caudal intraparietal area (CIP) receives simple, 3D visual data regarding object edges and surfaces from visual area V3A and selects object features suitable for grasping. Object related visual information then reaches the anterior intraparietal sulcus (AIP), which is

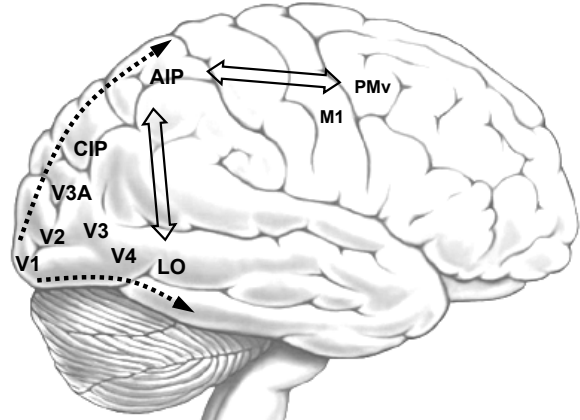


Fig. 1 Brain areas involved in vision-based grasping actions.

dedicated to analyze and select between grasping related visual features of objects in order to plan, and monitor the execution of, grasping actions [12]. Area AIP projects directly to (and receives backprojections from) the ventral premotor cortex PMv, which contains the motor primitives used to compose grasping actions. Finally, the primary motor cortex M1 is responsible for sending the proper activation signals to the muscles. Visual information flowing through the ventral pathway passes through V3 and V4 to reach the lateral occipital (LO) complex, the area of our cortex dedicated to object recognition [19]. The products of the visual analysis are thus, from the dorsal elaboration, precise information about position and geometry of the object and, from the ventral elaboration, data about expected weight, friction, and memories of previous interactions with the target object [10]. The detailed parameters of an action are completely determined by processing in the dorsal stream, and theoretically no contribution from the ventral stream is required. Nevertheless, it helps in the action selection process, providing semantic information regarding characteristics of the object and memories of past events [15].

When focusing on the use of visual information for action purposes, a key aspect to consider is the tight interconnection of primary visual with associative areas, which appears to interact through a series of feedforward and feedback connections. Indeed, visual perception is not a sequential process, but rather a distributed one, in which higher areas drive the job of primary ones in order to improve their visual knowledge in a goal-oriented way [7].

Attempts 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 [13], 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 [10], includ-

ing hypothesis on how the two streams of visual processing in our cortex coordinate and interact.

2.2 Robotics

In the engineering literature, the problem of selecting a set of contacts –a *grasp*– on an object for grasping it with a gripper is known as *grasp synthesis*, while the *grasp analysis* is the inverse problem, involving the evaluation of a given grasp [3]. A *grasp* is considered as the set of locations (points/regions) on the object surface where the fingers of the hand have to be placed for grasping the object. The grasp stability has often been evaluated in terms of *force* and *form closure* conditions, which ensure stability assuming point contacts with friction [3].

Most robotics researchers use a model of the object for the grasp search. For simplicity, some works have considered an object description that was restricted to lie on a plane [29]. Nevertheless, this description is unrealistic, since objects in the real world are intrinsically three-dimensional and in most cases are not well described by a 2D simplification. Works considering a 3D object description have normally used a model of the object, obtained or defined in an off-line stage [1, 23]; in other works [25], a grasping axis is computed using an octree-based reconstruction of the object. Alternatively, some authors use heuristics to reduce the number of candidates during the grasp synthesis and obtain a good grasp in short time [4]; 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 [26]. Quality measures for grasp analysis have been described in several works [5, 11].

Among vision-based works, in many cases there is a 2D grasp synthesis in one image, followed of a 3D reconstruction and/or validation [18]. 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 [33]. In general, the 3D reconstruction produces either a surface-based representation of the object [9] or a volumetric representation [24]. Nevertheless, the integration of this reconstruction with some task-oriented processing, such as grasp synthesis, has not been fully developed yet.

Pre-defined 3D object models have also been used in many works for the synthesis of grasps. In [14, 26] the decomposition of a 3D model into basic structural components is considered to find grasps on parts of the object that could be valid for the entire object. In other works, such as in [22],

the 3D object model is used to estimate the position and orientation of the object, and to position the robot with respect to the object, while the final grasp execution is guided by tactile sensor data. In fact, many works dealing not only with the synthesis but also with the execution of grasps make use of force or tactile sensor on the fingers of the robot hand. For instance, in [28], primitive force-based controllers are concurrently used to find an appropriate placement of the fingers on the object, and in [27] a robotic hand performs tactile exploration of the object, based on a set of exploratory primitives, in order to find a good grasp.

Finally, the development of a task such as the one considered in this paper has followed, in a classical approach, a *sense-plan-act* paradigm. This has been criticized for being too computationally expensive for allowing control in real time, and also for its lack of biological foundations. Behavior-based systems [2, 6] have proved their effectiveness in dealing with dynamic and complex environments, and have been widely used in many fields of robotics. Moreover, although this paradigm has been more widely used in other areas of robotics, it is being increasingly used in the development of manipulation tasks [34].

3 The filter-based architecture and its application to vision-based grasping

In this section, an architecture is proposed for the adaptation of neuroscience models to a robotic setup. A framework for the implementation of a model of vision-based grasping based on the early separation and late integration of visual analysis through the two visual streams is presented.

3.1 The filter-based architecture (FBA)

The proposed architecture, intended for its application on artificial devices, has some intrinsic limitations with respect to biological models, which influences aspects such as the degree of parallelism and the flow of information. It uses the basic types of components shown in Figure 2 [31]:

- *Virtual sensors*. Components that provide data acquired from real sensors 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

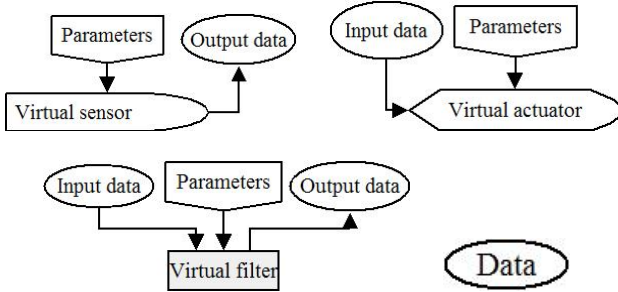


Fig. 2 Basic components of the FBA architecture.

a specific cortical area. In the brain, they are usually 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 produced and processed by the above modules. They represent a simplification of the information flow connecting brain areas.

Each output data set of a virtual sensor or filter can constitute the input of a virtual actuator or another filter.

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.

In addition to its functionality, it is the set of interfaces of a component what characterizes it. This allows to group a set of components into a single unit that can be considered as a higher level sensor, filter or actuator. Additionally, filters can internally implement the different stages in the execution of a task. Transitions between these stages may depend on the input data, the filter output, or a combination of both.

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.

We believe that our architecture is especially suitable for bridging the gap between neuroscience models and robotic applications. On the one hand, it offers the modularity and flexibility required for modeling brain functions and data-streams. On the other hand, the rigorous and formal representation of these data-streams allows for a further abstraction that is necessary in order to translate a computational skill into a practical implementation of a real-world task. Figure 3, which is explained in more detail in the following

sections, shows how the components of this architecture can be composed to form a relatively complex task framework, in our case dedicated to vision-based grasp planning.

3.2 The vision-based grasping model

This section considers the development of a model which includes the neuroscience findings described in Section 2 within the framework of the filter-based architecture. According to the FBA, the simplified model of the task performed by the two cortical visual streams in the context of vision-guided grasping actions is depicted in Figure 3, using the blocks of Figure 2. The data flow through the two visual streams, as explained in 2.1, is shown. Along the dorsal stream, grasp relevant visual information is processed and refined in order to generate and evaluate possible grasp configurations. Perceptual visual elaboration, relying upon more abstract features such as color and overall shape, is performed along the ventral stream, which will aid the final grasp selection through access to previous experience.

From a practical point of view, the main novelty of applying such a framework to a robotic setup would be the parallel computation performable 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. The on-line data gathered and processed by the dorsal stream is complemented and enriched by the previously stored knowledge recovered by the ventral stream. Each pathway thus contributes to the final behavior with the most important results of its elaboration. Moreover, both streams interact with each other, the ventral helping the dorsal in action selection, and the dorsal providing spatial visual information that can be used to accelerate and improve object classification and recognition.

Besides the separation of the two streams, a critical aspect is the recurrent connectivity between sensory and associative areas. On the one hand, the link between visual areas (V1, V2) and dorsal associative cortex (mainly CIP, AIP and MT, not shown) represents the gradual enrichment of visual information from basic features such as blobs or edges, to the elaborated representation of different kind of surfaces and complex shapes. On the other hand, such link serves the function of providing additional details to current representations, being the refinement driven by the needs of the dorsal stream areas which select the object or object features that deserve more attention.

In the next section, we provide a first implementation of part of the described model. 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 we refine and complement the available data, until

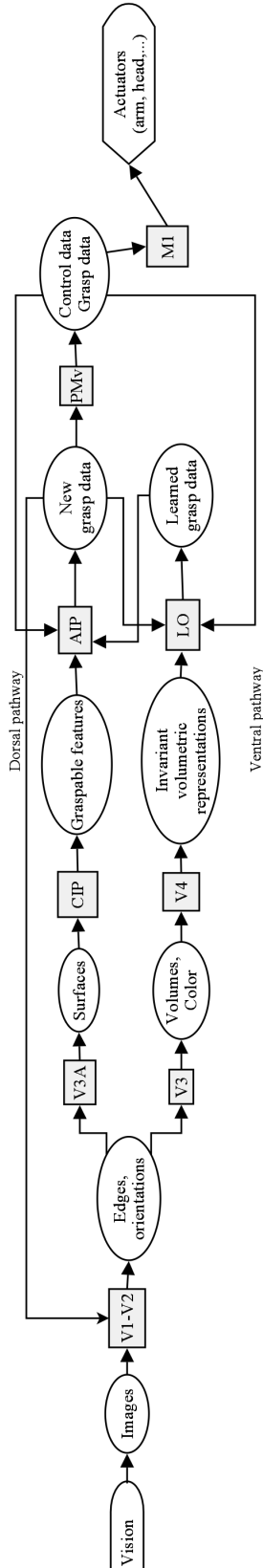


Fig. 3 Development of the vision-based grasping model using the FBA architecture.

they are considered reliable enough to proceed to the synthesis of an executable grasp.

Later on, we plan to include in the system the contribution of the ventral stream. The idea is that memory of previously executed grasping actions on given objects can be recalled and facilitate the analysis/selection of grips. The input of the recognition step is information on the shape, the color and texture of an object. Identifying the object will translate in a classification task, to categorize the target in one of some known object classes. Outputs will thus be the object identity, and its composition, which in turn allows to estimate its weight distribution and the roughness of its surface, that are valuable information at the moment of planning the action. Recognition is not a true/false process, and a reliability index of the extracted information needs to be provided, as the classification could be more or less certain: if it is considered very unreliable, more importance will have the on-line visual information gathered by the dorsal stream. Beside the recovery of memorized object properties, the recognition allows to access memories of previous grasping experiences. These can be used to associate an object with basic natural grips, to recall the outcome of old actions on that object, and this information is used to bias the grasp selection.

4 Modeling grasp-synthesis

Considering the model described in Section 3.2, we have focused on one specific task, the on-line grasp visual synthesis performed by the dorsal stream in conjunction with the visual areas, through visual exploration of target objects.

4.1 The grasp-synthesis task

We develop the grasp-synthesis strategy so that it is based on the visual exploration of the object. The proposed exploration is 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 experimental setup considered in this work, shown in Figure 4, 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 their contour can be extracted by the vision system in each acquired image. 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 5 provides a general description of this task, which is composed of the following stages:

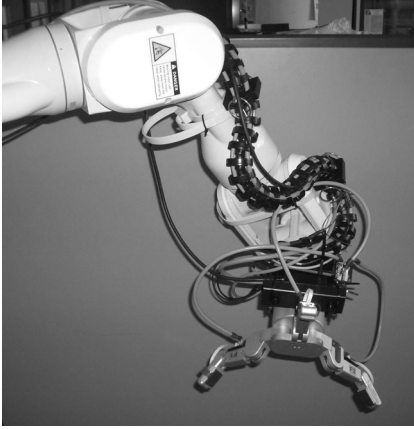


Fig. 4 Experimental setup for the visual object exploration.

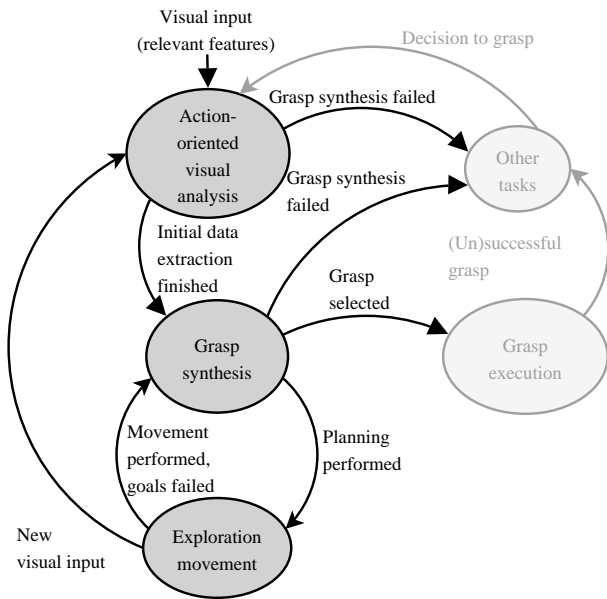


Fig. 5 Description of the grasp-synthesis task.

- *Action-oriented 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 any case, the grasp-synthesis can be preceded or followed by other tasks.

Due to the inherent complexity of the task and our special interest in visual elaboration, we focus in the following sections on the vision analysis issues from Figure 5 associated to the grasp synthesis only, rather than on the control and planning of the exploratory movement.

4.2 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 CIP. 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-premotor loop comes into action) or it has not been possible to perform it (for example for a sudden occlusion).

Our system builds at this stage a rough 3D object representation, which will be refined in the *Grasp Synthesis* stage to identify possible grasp zones. The object representation is based on the use of an octree structure, which allows to control the degree of detail of the representation. An octree structure is based on a node (*voxel*) that can be recursively decomposed in eight children, where each node represents a cube in physical space and each child an octant of its parent. The octree can be *carved* in order to remove nodes corresponding to space not belonging to the object [32].

The setup data required at this stage is 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.

The carving of the octree, in order to build an object model, is performed following algorithm 1 using the contours of the selected object, extracted by the modules in charge of basic image processing (corresponding to the primary visual cortex). The octree is projected onto the images provided by the vision system and then compared with the contours of the selected object. This is used to remove (*carve*) voxels that, based on this projection, are outside the space occupied by the object. The images correspond to different views of the object, obtained during an initial exploratory movement of the robot around it. They are binarized and contours extracted from them. The selection of the contours corresponding to the target object is based on generic criteria, such as the size and the position in the image.

The above algorithm is applied to each input contour-based object description. The level of detail of this reconstruction is controlled by limiting the number of levels in the octree and the size of the octree nodes. The decision in algorithm 1 on whether the carving of an octree node is required

or not is determined by algorithm 2. This algorithm is designed to use for this purpose generic criteria related to the overlapping between the object and the octree node, as well as grasp-specific criteria. At the *visual analysis* stage, only the generic criterion, which considers as a non-null occupancy of the projection of a voxel on the object, is used.

Algorithm 1 Carve 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
      else if node projection is outside the object then
        Mark node as empty
      end if
    end for
  end if
end for

```

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 size  $\geq$  threshold then
  if overlapping between object and node projection then
    if voxel occupancy  $\geq$  threshold then
      carving required  $\leftarrow$  true
    else if carving required based on grasping criteria then
      carving required  $\leftarrow$  true
    end if
  end if
end if

```

The octree-based representation obtained at this stage is used as just a rough description of the object, which will be later, at following stages, refined with higher level of detail in order to highlight the zones more interesting for grasping.

4.3 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 grasp synthesis, it is checked if a reliable grip can be produced; if so, the control passes to the motor areas in charge of organizing the target movements; otherwise, a new exploratory movement is performed and new visual data gathered and added to that already available.

Once the rough, initial octree-based object description has been obtained, a refinement of this octree is performed

in order to highlight, through a more detailed modelization, the regions more interesting for grasping.

Algorithm 3 provides an outline of the behavior at this stage. Essentially, it analyzes the object description stored and produced during its exploration. In case more information about the object is required for the evaluation of all the criteria considered for the grasp synthesis, a new exploratory movement is planned accordingly, based on the available object data [30], taking into account the object regions more interesting for grasping and in order to provide more detailed information about them.

If the available information is enough to perform the above evaluation, we use a number of heuristic criteria that we have previously defined and validated experimentally [11] to decide if a reliable grip can be extracted from it. Some of the criteria are related to the hand kinematics, but in this case we want to keep the process purely visual, and as flexible as possible. Thus, only those criteria directly related to the shape, size, and disposition of the grasp regions are taken into account. More exactly, the ones that are considered more useful and easily adaptable to the new situation are: the symmetry in the distribution of the grasp regions (criterion S1 in the original publication); the safety margin provided by a given set of graspable features (criterion S3); and the curvature of the regions (slight concavities usually guarantee more stable grips, criterion S4). Possible sets of regions representing a grip are thus evaluated using the normalized versions of the criteria, and if a candidate grip reaches a given reliability threshold, then the exploration can terminate. Future work will also consider the selection of a corresponding motor action, out of a basic vocabulary of actions, and a grasp preshape for the hand [20], which is now manually defined by the operator.

Algorithm 3 Exploration-based grasp synthesis

```

Analyze the object information
if more information is required then
  Plan a new exploratory movement
  Trigger event Planning performed
else
  Attempt grasp selection from available object information
  if grasp selection has been successful then
    Select a motor action and a pregrasp
    Trigger event Grasp selected
  else
    Trigger event Grasp synthesis failed
  end if
end if

```

4.4 Exploration movement

The activity in this stage is oriented to the extraction of object information during the execution of a planned explo-

ration movement. This information is used to incrementally update the object description produced by the visual analysis stage. Algorithms 1 and 2 are used also at this stage for the refinement of the octree. Both the generic and the grasp-specific criteria are considered in the *exploration movement* stage for deciding on the carving of the voxels; as the grasp-specific criterion, we consider the curvature of the object contour below a given threshold.

The sequence of activities in this case can be briefly described as follows:

- *Visual assessment of available feature data.* The system, in its aim of generating possible grips, assesses the quality of the available data through the use of visual criteria [11]; 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. Therefore, the decision on whether to carve a voxel or not in algorithm 2 is related not only to the voxel occupancy, but also to the fulfillment of the grasping criteria.
- *Movement control.* The movement planned in the *grasp synthesis* stage is executed. In the human 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 motor cortex responsible for the joint movement of sensor and effector.
- *Enrichment of visual knowledge.* Further 3D visual data, regarding those portions of the object surface assessed as more appropriate for grasping, is gathered and added to the information already available in the dorsal areas (CIP-AIP), which can repeat the synthesis-exploration cycle until a reliable grasp configuration is found.

5 Results

This section shows the results of the visual analysis described in Section 4 using the experimental setup introduced in Section 4.1. They correspond to the visual processing performed in the *action-oriented visual analysis*, *grasp synthesis* and *exploration movement* stages shown in Figure 5. As indicated in this figure, the first of these stages uses as input the relevant features extracted from the images acquired by the camera mounted on the robot arm during its initial and exploratory movements around the object; in particular, we have considered as relevant feature the contour of the object.

The images acquired during the exploratory movement of the robot arm are processed according to the procedures described in Section 4. Figure 6 shows the results of the *visual analysis* stage based on the initial views of a selected object. Using an octree structure, a rough modeling is per-

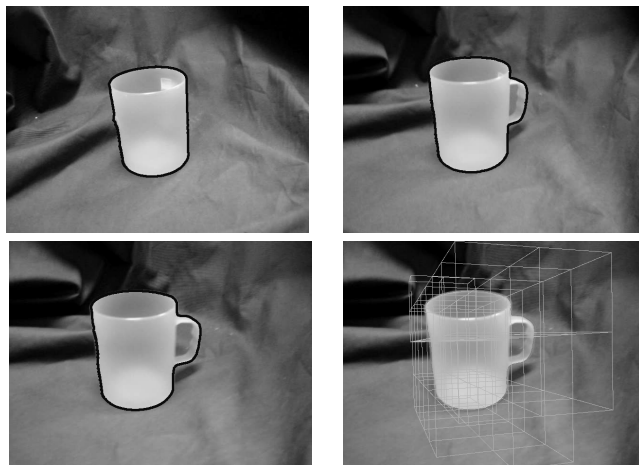


Fig. 6 Initial data extraction. Sequence of object views, with extraction of relevant features, followed (bottom-right image) with a rough object reconstruction.

formed that will be used afterward for the grasp-based object exploration. The generated octree has a maximum of four levels of depth. Three images from different viewpoints have been used. They are just slightly different, taken only from a few centimeters away, but enough for providing a stereoscopic effect. 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, in which the object is observed from different views. Some of these views for a sample object are provided in figure 7; external views corresponding to this exploratory movement can be observed in figure 8. As indicated in section 4, a generic refinement is performed first in order to better define the shape of the object. Then, more specific criteria intervene so that the analysis focuses on those areas of the object that are more interesting for grasping. The regions of the object that, based on the above criteria, are considered less suitable are ignored or observed with less detail. Figure 9 shows the process of applying this refinement to the rough modeling given in figure 6, with figure 9d corresponding to the final refinement. Further examples of this process are provided in figure 10.

The grasp is selected according to visual criteria [11] on the regions of the object that have received more attention during the reconstruction process. Figure 11 shows some examples of grasps executed on these regions.

6 Conclusions

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 compu-

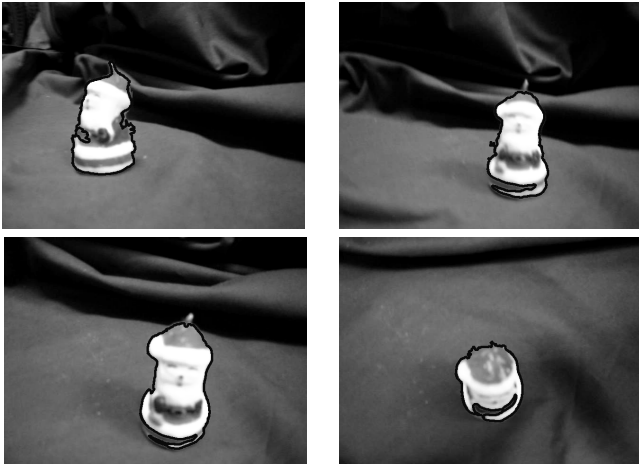


Fig. 7 Different views of an object during the exploration movement.

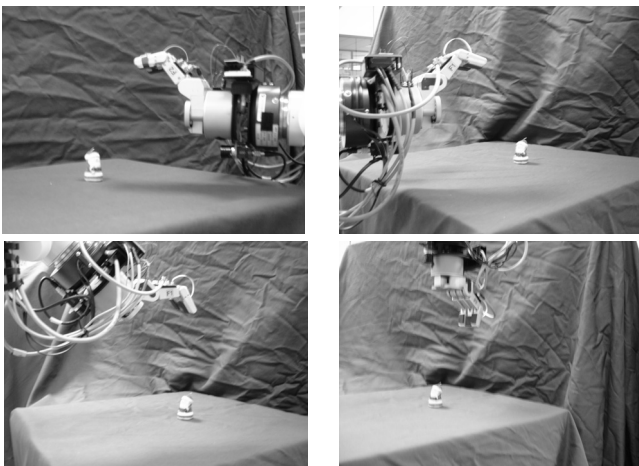


Fig. 8 External views of the exploration movement around an object.

tations 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 visual analysis of the object, future work will have to consider the cases in which it is difficult to extract relevant information from low-resolution data in the earlier stages of the grasp search. As a first, major improvement, we plan to extend the grasp-oriented visual analysis to the case in which experience of previous grasping actions and object knowledge is available (contribution of the ventral stream). This should not be done in a dichotomic way, but rather as a modulation of the influence of the two streams: a higher confidence in the object recognition/classification process reflects in a stronger influence of past grasping experiences, whilst a more uncertain recognition leads to a more exploratory behavior, giving more importance to actual observation.

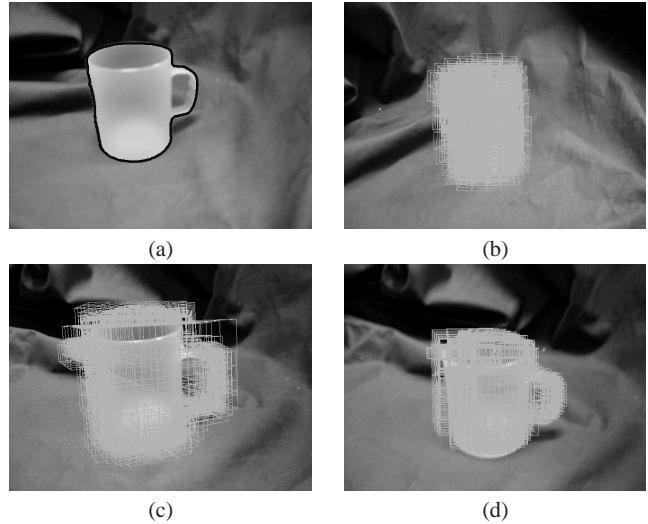


Fig. 9 Exploration movement for grasp synthesis. View after the initial object reconstruction (a), which triggers a generic reconstruction (b)-(c), followed by a grasp-specific reconstruction (d).

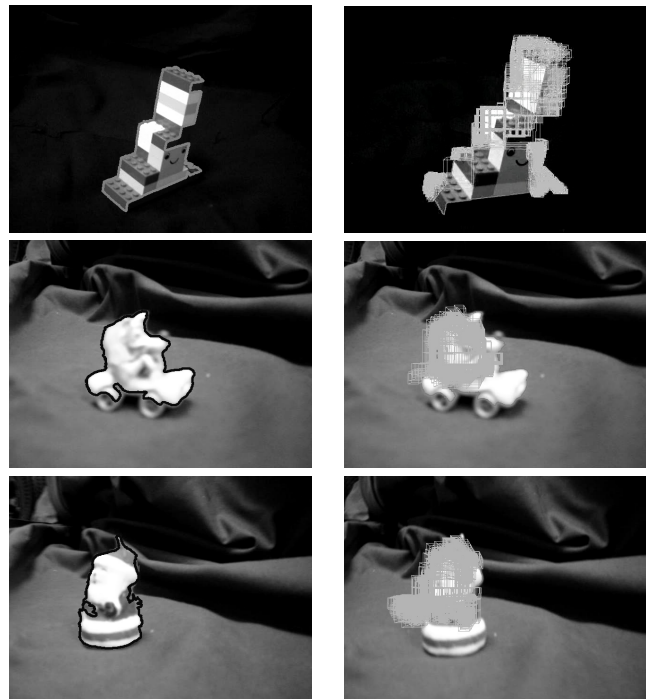


Fig. 10 Other results of object reconstruction for grasp synthesis.

A second important development of our work is toward a safer and more efficient action execution. It is long known that visual information alone is hardly enough to provide stability to real world grips [21], and recent findings [17] show that the parietal cortex, and namely AIP, is involved in a cross-modal analysis of object features which involves tactile information. The practical goal to pursue is thus to integrate in the analysis of graspable features data regarding the contact between object and effector.

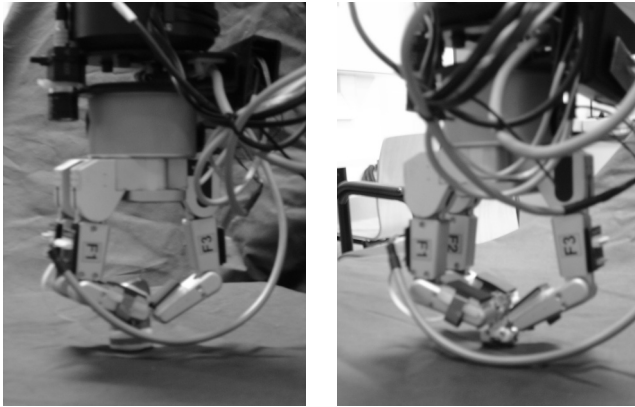


Fig. 11 Examples of grasp execution on the objects from Figure 10.

Finally, the framework described in this paper is oriented to be used in a relatively autonomous system, which would have to handle on its own the execution of specified tasks with 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. 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*, pages 2124–2130, Orlando (Florida), USA, May 2006.
2. R.C. Arkin. *Behavior-Based Robotics*. Intelligent Robotics and Autonomous Agents series. The MIT Press, Cambridge (Massachusetts), USA, 1998.
3. A. Bicchi and V. Kumar. Robotic grasping and contact: A review. In *Proc. IEEE Intl. Conf. on Robotics and Automation*, pages 348–353, April 2000.
4. 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*, pages 1890–1896, Detroit (Michigan), USA, May 1999.
5. 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*, pages 319–325, New Orleans (Louisiana), USA, April 2004.
6. R. Brooks. *Cambrian intelligence: The early history of the new AI*. MIT Press, Cambridge (Massachusetts), USA, 1999.
7. J. Bullier. Integrated model of visual processing. *Brain Res Brain Res Rev*, 36(2-3):96–107, October 2001.
8. U. Castiello. The neuroscience of grasping. *Nat Rev Neurosci*, 6(9):726–736, September 2005.
9. F. Chaumette, S. Boukir, P. Bouthemy, and D. Juvin. Structure from controlled motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(5):492–504, May 1996.
10. E. Chinellato, Y. Demiris, and A.P. del Pobil. Studying the human visual cortex for achieving action-perception coordination with robots. In *IASTED Intl. Conf. on Artificial Intelligence and Soft Computing (ASC 2006)*, 2006.
11. E. Chinellato, A. Morales, R.B. Fisher, and A.P. del Pobil. Visual features for characterizing robot grasp quality. *IEEE Transactions on Systems, Man and Cybernetics, Part C*, 35(1):30–41, February 2005.
12. J. Culham. Human brain imaging reveals a parietal area specialized for grasping. In N. Kanwisher and J. Duncan, editors, *Functional Neuroimaging of Visual Cognition: Attention and Performance XX*, pages 417–438. Oxford University Press, 2004.
13. A.H. Fagg and M.A. Arbib. Modeling parietal-premotor interactions in primate control of grasping. *Neural Networks*, 11(7–8):1277–1303, October 1998.
14. C. Goldfeder, P. Allen, R. Pelossof, and C. Lackner. Grasp planning via decomposition trees. In *Proc. IEEE Intl. Conf. on Robotics and Automation*, pages 4679–4684, Rome, Italy, April 2007.
15. M. Goodale and D. Milner. *Sight Unseen*. Oxford University Press, 2004.
16. M. A. Goodale and A. D. Milner. Separate visual pathways for perception and action. *Trends Neurosci*, 15(1):20–25, January 1992.
17. C. Grefkes, P.H. Weiss, K. Zilles, and G.R. Fink. Crossmodal processing of object features in human anterior intraparietal cortex: an fMRI study implies equivalencies between humans and monkeys. *Neuron*, 35(1):173–184, Jul 2002.
18. 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*, pages 272–278, Kyongju, Korea, 1999.
19. T.W. James, G.K. Humphrey, J.S. Gati, R.S. Menon, and M.A. Goodale. Differential effects of viewpoint on object-driven activation in dorsal and ventral streams. *Neuron*, 35(4):793–801, August 2002.
20. M. Jeannerod, M.A. Arbib, G. Rizzolatti, and H. Sakata. Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends in Neuroscience*, 18(7):314–320, 1995.
21. R. S. Johansson and G. Westling. Roles of glabrous skin receptors and sensorimotor memory in automatic control of precision grip when lifting rougher or more slippery objects. *Exp Brain Res*, 56(3):550–564, 1984.
22. D. Kragić, S. Crinier, D. Brunn, and H.I. Christensen. Vision and tactile sensing for real world tasks. In *Proc. IEEE Intl. Conf. on Robotics and Automation*, pages 1545–1550, Taipei, Taiwan, September 2003.
23. E. Lopez-Damian, D. Sidobre, and R. Alami. Grasp planning for non-convex objects. In *Proc. of the 36th International Symposium on Robotics*, Tokyo, Japan, November 2005.
24. P.R.S. Mendonça, K.-Y.K. Wong, and R. Cipolla. Epipolar geometry from profiles under circular motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):604–616, June 2001.
25. C. Michel, V. Perdereau, and M. Drouin. Extraction of the natural grasping axis from arbitrary 3D envelopes provided by voxel coloring. In *Proc. of the 36th International Symposium on Robotics*, Tokyo, Japan, November 2005.
26. A.T. Miller, S. Knoop, H.I. Christensen, and P.K. Allen. Automatic grasp planning using shape primitives. In *Proc. IEEE Intl. Conf. on Robotics and Automation*, pages 1824–1829, Taipei, Taiwan, September 2003.
27. L. Natale and E. Torres-Jara. A sensitive approach to grasping. In *Proc. Sixth Intl. Conf. on Epigenetic Robotics*, Paris, France, September 2006.
28. R. Platt, A.H. Fagg, and R.A. Grupen. Nullspace composition of control laws for grasping. In *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, pages 1717–1723, Lausanne, Switzerland, October 2002.
29. J. Ponce, D. Stam, and B. Faverjon. On computing force-closure grasps of curved two dimensional objects. *The Intl. J. of Robotics Research*, 12(3):263–273, June 1993.
30. G. Recatalá, E. Chinellato, A.P. del Pobil, Y. Mezouar, and P. Martinet. 3D grasp synthesis based on a visual cortex model. In *Proc. of the First IEEE / RAS-EMBS Intl. Conf. on Biomedical Robotics and Biomechanics*, Pisa, Italy, February 2006.

31. G. Recatalá, P.J. Sanz, E. Cervera, and A.P. del Pobil. Visual servoing for the tool-to-object positioning with respect to welding points. In P. Drews, editor, *Mechatronics & Robotics'04*, volume III, pages 1241–1246, Aachen, Germany, September 2004. Sascha Eysoldt Verlag.
32. M.A. Yerry, and M.S. Shepard. A modified quadtree approach to finite element mesh generation. *IEEE Computer Graphics and Applications*, 3(1):39–46, January 1983.
33. G. Zeng, S. Paris, M. Lhuillier, and L. Quan. Study of volumetric methods for face reconstruction. In *IEEE Intelligent Automation Conference*, 2003.
34. R. Zöllner, M. Pardowitz, S. Knoop, and R. Dillmann. Towards cognitive robots: Building hierarchical task representations of manipulators from human demonstration. In *Proc. IEEE Intl. Conf. on Robotics and Automation*, pages 1547–1552, Barcelona, Spain, April 2005.



Gabriel Recatalá holds a B.Sc. degree (1996) and a Ph.D. (2003) in Computer Science Engineering from the “Jaume I” University (Spain). He has been appointed as Visiting Scientist at the Department of Computer Science of the University of Minnesota (1998, USA), the Institute for Real-Time Computer Systems of the Technical University of Munich (1999, 2000, 2001, Germany), and the LASMEA laboratory of the “Blaise Pascal” University in Clermont-Ferrand (2002, France). He is teaching assistant at the “Jaume I” University since 1999. His current research interests include robotics, computer vision, visually-guided grasping, visual servoing, and behavior-based architectures.



Eris Chinellato received his MSc in Artificial Intelligence, together with the Best Student Prize, from the University of Edinburgh (UK) in 2002, and his Industrial Engineering Degree from the Università degli Studi di Padova (Italy) in 1999. He is now pursuing his PhD in the Robotic Intelligence Lab of the Jaume I University (Spain). His interdisciplinary research is mainly focused, but not restricted to, the use of visual information for grasping actions in natural and artificial systems. He has published in influential journals and proceedings in robotics, neuroscience, and computational neuroscience. He has served as reviewer and program committee member for international journals and conferences, and collaborated with renowned scientists such as M.A. Goodale and R.B. Fisher.



Ángel P. del Pobil is Professor of Computer Science and Artificial Intelligence at Jaume I University (Spain), and founder director of the Robotic Intelligence Laboratory. He holds a B.S. in Physics (Electronics, 1986) and a Ph.D. in Engineering (Robotics, 1991), both from the University of Navarra (Spain). His Ph.D. Thesis was the winner of the 1992 National Award of the Spanish Royal Academy of Doctors. Prof. del Pobil is author or co-author of over 120 research publications, and has been invited speaker of 34 tutorials, plenary talks, and seminars. His past and present research interests include: motion planning, visually-guided grasping, service robotics, mobile manipulators, visual servoing, learning for sensor-based manipulation, and the interplay between neurobiology and robotics.



Youcef Mezouar received the Ph.D degree in computer science from the University of Rennes 1, Rennes (France), in 2001. He has been Postdoctoral Associate in the Robotics Laboratory, Computer Science Department, Columbia University, New York, NY, and currently holds a position as Associate Professor at the Blaise Pascal University (France). His research interests include robotics, micro-robotics, computer vision, and vision-based control.



Philippe Martinet graduated from the CUST, Clermont-Ferrand (France), in 1985 and received the Ph.D. degree in electronics science from the Blaise Pascal University, Clermont-Ferrand (France), in 1987. From 1990 to 2000, he was assistant Professor at the CUST in the Electrical Engineering Department, Clermont-Ferrand. Since 2000, he has been a Professor at the Institut Français de Mécanique Avancée (IFMA), Clermont-Ferrand. He was the leader of the group GRAVIR (over 74 people) from 2001 to 2006. He is performing research at the Robotics and Vision Group of LASMEA-CNRS, Clermont-Ferrand. (Korea). His research interests include visual servoing, force-vision coupling, multi-sensor based control, autonomous vehicles, and modeling and control of complex machines.