3D Grasp Synthesis Based on a Visual Cortex Model

Eris Chinellato, Gabriel Recatala Ballester, Angel P del Pobil, Youcef Mezouar, Philippe Martinet

3D Grasp Synthesis Based on a Visual Cortex Model

Gabriel Recatalá, Eris Chinellato and
Angel P. del Pobil
Robotic Intelligence Lab
Dept of Computer Science and Engineering
Jaume I University
8029AP Castellón, Spain
grecata,eris,pobil@icc.uji.es

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

***Abstract*— In this paper, the problem of object grasping is considered from both a biological and an engineering point of view. A model of information processing for the grasp synthesis and execution is described based on recent findings from neuroscience. Taking into account the differences between robotic and biological systems, this paper proposes the adaptation of that model to the peculiarities of a robotic system, instead of mimicking it. For this purpose, an architecture is proposed that allows the scalability of this model and its integration within more complex tasks. The grasp synthesis is designed as integrated within the extraction of a 3D object description, so that the object reconstruction is driven by the needs of the grasp synthesis. The integration is formulated as a framework where different grasp synthesis strategies could be applied.**

***Index Terms*— active perception, models of human manipulation, models of human perceptual systems, neuroscience models in robotics, perception and action**

I. Introduction

The ability to manipulate every kind of objects in a dexterous way has been widely analyzed from both a robotic and a neurophisiological point of view. Nevertheless, there are still important differences between humans and robots that influence the way robotic manipulation applications can be defined. In the first place, human hands are characterized for having five soft fingers and a high degree of dexterity, whereas most robotic hands feature a low level of dexterity, and have a more reduced number of fingers, with a hard surface. 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, globaly, more degrees of freedom than current robotic setups. 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 taylored, 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.

This paper approaches the grasp search problem from a biological point of view. In particular, it proposes the adaptation of a model of information processing for vision-based grasping in the human brain to a robotic system. An architecture is proposed for the development of the above model, following behavior-based guidelines. This architecture supports the nesting and the concatenation of processing modules in a structured way. In addition, the processing modules can have states, so that each module behaves individually as an automata, and, thanks to the nesting of modules, automatas can run at different levels of abstraction within the whole system.

The grasp search we propose is formulated through the integration of object reconstruction and grasp search procedures. This search is based on the object exploration, with active vision, for an incremental 3D object reconstruction, driven by the search of features that characterize the grasp configuration that has to be computed. This reconstruction increases its level of detail on the regions of the object where those features may be found. The proposed procedure is defined as a framework where different grasp search and analysis strategies can be used.

II. Neuroscience and robotics background

One of the main concerns in neuroscience has been the study of the areas of the brain involved in the different stages of a manipulation task, as well as of the flow of information through these areas [1]. This path of research has been followed both on humans and other primates. In our work, we mainly refer to the former, unless stated otherwise. Visual data in humans flows from the retina to the lateral geniculate nucleus (LGN) of 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 V1 to different association areas, the posterior parietal cortex (PPC) and the inferior temporal (IT) cortex (Figure 1).

The traditional distinction [2] talks about ventral ”what” and dorsal ”where/how” visual pathways. Object informa-

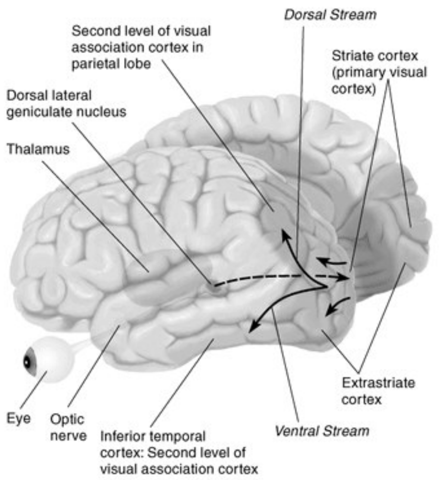

Fig. 1. Dorsal and ventral visual pathways.

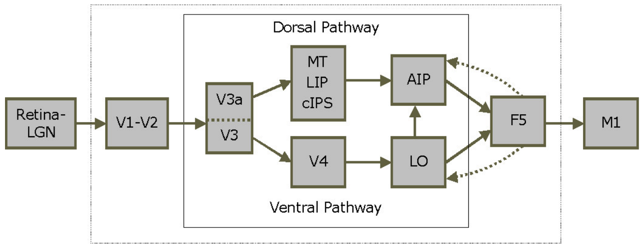

Fig. 2. Block diagram of a grasping system based on human physiology.

tion flowing through the ventral pathway passes through V2 and V4 to the lateral occipital (LO) complex, which is related with object recognition [3]. Through the dorsal pathway, object-related visual information reaches area AIP in the intraparietal sulcus, which is concerned with analyzing visual features in order to organize grasping actions. Area AIP projects mainly to area F5 of the premotor cortex, which contains the motion primitives used to compose grasping actions, and receives most of its input, through other areas of the intraparietal sulcus (mainly the lateral part LIP and the caudal part cIPS), from the extrastriate cortex V3a, which seems to be responsive to stimuli orientation as well as to motion and color.

Although recent studies confirm that the dorsal stream is more oriented to action-based vision, whilst the ventral one is more suitable to categorization [4], such distinction is not completely sharp. New findings [5] suggest that both streams are involved in most vision-related tasks, only in a different way and degree according to the nature of the task. It is also largely accepted that the two streams are strictly related, but how they communicate is still mostly unknown. One of the most complete models of the information flow during grasping is the FARS model [6], which focuses on the final part of the process, more related with action-execution. Nevertheless, no robotic applications have been yet developed following this path and the integration between both streams is nearly unexplored [7]. An additional model is described in [8] that considers the integration of the visual information processed along these pathways.

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 [9]. Other authors have considered partially-restraining grasps, which allow some degrees of freedom to the object [10]. Alternatively, some authors use heuristics to reduce the number of candidates during the grasp synthesis and obtain a good grasp in short time [11]; 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 [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. 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. In general, the 3D reconstruction produces either a surface-based representation of the object [13] or a volumetric representation [14].

III. THE VISION-BASED GRASPING MODEL

In this section, a model of the processing of visual information in human grasping is described, based on recent studies from neuroscience. This model is developed in further sections in order to taylor it to a robotic system.

In our framework, outlined in figure 2, the visual input is processed in two parallel ways, one more concerned with perceptive information about the object nature, the other 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 labelled after the name of the brain regions their behavior is associated to. In the diagram of figure 2 there should be many more connections, and all arrows should be bidirectional, 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.

In our model, the visual blocks V1 and V2 provide as output basic features, mainly edges, corners, or simple contours. They are used by V3 to reconstruct more complex ones. As it can be observed in figure 2, areas V3 and V3a appear slightly separate, as their processing is not exactly the same. In fact, the visual analysis is likely to be performed in different ways by the two pathways starting from V3. If on the one hand achieving a quick and reliable object recognition is probably better done through a volumetric analysis, it seems more plausible for the dorsal stream to look for appropriate surfaces (and not volumes) on which to put the fingers [15]. Thus, V3a can use the basic features obtained by V1-V2 to reconstruct surfaces, while V3 can search for volumetric features.

Surface information coming from V3a is the input of areas MT, LIP and cIPS. In humans, MT is more concerned about movements, while LIP seems to be responsible of storing and remapping visual memory in eye-centered coordinates [7]. Brain area cIPS is believed to code surface orientations [16] and also object affordances, as it does not recognize the same object seen from two different view points [3]. In our model, cIPS is especially critical, as it is the block in which the search for graspable surfaces is carried out. The possible grasping regions found by cIPS are used by AIP, the main grasping area, responsible for finding the candidate grips joining appropriate surfaces.

On the ventral side, V4 uses the output of V3 to build a viewpoint-invariant simple reconstruction of the object, using volumetric primitives (see e.g. [12]). Then, LO merges spatial and color data with stored information about previously observed objects to recognize the target and access its memory on how it has been grasped before. Regarding the last point, it is still unknown if this is what actually happens in our brain, and our solution so far is not supported by neuroscience findings. Also, LO most likely projects to F5 only indirectly, through the prefrontal cortex, which surely plays a role in the process (also involving the *task* which is to be performed through the action, an aspect that we decided not to treat at this point of our study).

The output from AIP and LO is finally used by F5 to choose the most appropriate grip for that object in that situation. Although we do not know exactly in which way the two pathways interact, our idea is that the ventral information can intervene in the grip generation (AIP) and selection (F5) process inhibiting unsuitable grips, or families of grips (e.g. precision grips instead of power ones). The power of the intervention can be modulated according to the degree of confidence of the object recognition process, leaving more or less influence on the dorsal analysis. After the selection process is done, when AIP receives feedback from F5 telling which grip will be finally performed, it forwards the information to LO, in order to memorize the selected action for future reference.

IV. The filter-based architecture (FBA)

In this section, an architecture is proposed that supports the development requirements of the model described in the previous section. This architecture, however, is intended for its application on a computer, so it has some intrinsic limitations with respect to the biological model, which influences aspects such as the degree of parallelism and the flow of information. This architecture uses the following basic types of components, shown in figure 3:

- *Virtual sensors.* Components that provide data acquired from real sensors installed in the system (cameras, infrared cells, etc.). They correspond more to the primary sensory areas of the human cortex than to the sensory organs. In our vision case, the output of a camera is more similar to that of LGN or V1 than that of the retina.
- *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, as M1 in our case.
- *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 associative 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. It represents an over-simplification of the information flow connecting brain areas.

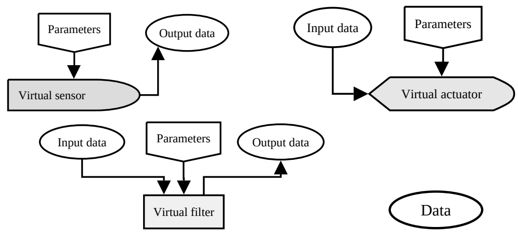

Fig. 3. Basic components of the FBA architecture.

Virtual sensors and actuators are connected through a chain of filters, in which the input/output data sets describe the flow of information along the chain. 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 consitute an internal, non-centralized memory spread along the chain.

Virtual sensors, actuators and filters have interfaces, through which they are interconnected. As shown in figure 3, 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 as a sensor, a filter or an actuator. 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 be in different states, performing a different processing in each state. Transitions between states may depend on the input

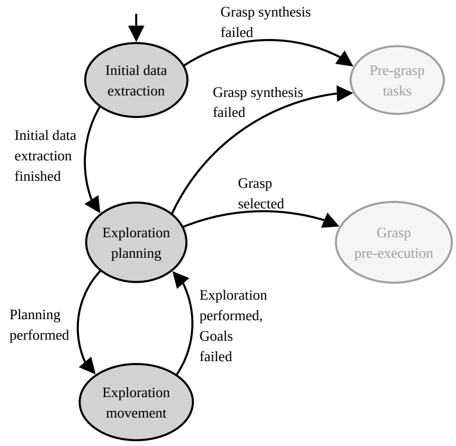

Fig. 4. Automata-based description of the grasp-synthesis task.

data, the filter output, or a combination of both. Figure 7 provides an outline of the model described in section III developed using the FBA architecture.

V. The grasp-synthesis model

A. The grasp-synthesis task

This section develops the model described in section III. There are, however, some aspects of this model, such as the processes of object reconstruction and grasp selection, that have not been yet fully explained from a neuroscientific point of view. Taking this into account, we develop the grasp-synthesis strategy so that it is based on the visual exploration of the object. This exploration is active, guided by the need of searching or computing specific data that are required for the grasp synthesis. This strategy is described from a general point of view, since the purpose of this paper is not to provide a detailed description of it; instead, it is given as a framework within which different grasp synthesis and analysis criteria could be tested.

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

- *Initial data extraction*. Initial stage, in charge of gathering a set of data to start the object exploration.
- *Exploration planning*. This is an explotation 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.

Due to the inherent complexity of the model described in section III, we have restricted ourselves in this paper to the components more related to the grasp synthesis that would be more active in these stages. Therefore, some areas of that model, such as MT and LIP, less involved in this process, have not been considered in this development.

The following sections describe in more detail the exploration planning and the exploratory movement.

B. Exploration movement

Although a planned movement is executed in this stage, the gathering of information along the dorsal pathway is determined by a set of quality criteria [17]. These criteria are associated to the selection of *affordances*, or *grasp surfaces*, which are portions of the object surface that are considered appropriate for grasping. The specification of these criteria is part of the grasp synthesis strategy and is left open in this paper, in order to allow the development of different strategies and support different robotic setups.

Figure 5 provides a filter-based outline of this stage. The overall behavior can be briefly described as follows:

- *Regarding the primary visual perception.* Extraction of basic features for driving the exploratory movement, as well as others that can be used later for building a model of the object (V1-V2).
- *Along the dorsal pathway.* Reconstruction of portions of the object surface (V3a) and search, within them, of grasp surfaces (cIPS). Such surfaces may be priorized according to the degree of fulfillment of the selected quality criteria.
- *Along the ventral pathway.* Search of volumetric features (V3) and 3D reconstruction based on volumetric primitives (V4).
- *Movement control.* Control of the exploratory movement (F5). The goals of this movement have been stored during the planning stage in an internal memory of this module.

C. Exploration planning

This is the stage in which the grasp synthesis is actually performed, using the initial data set and additional information collected during the object exploration. It is therefore a deliberative stage. In this stage, the grasp surfaces are analyzed along the dorsal pathway in order to select *compatible surface sets*, which are combinations of such surfaces that are thought to be appropriate for grasping. Using this information and the one provided through the ventral pathway, a grasp is selected based on a set of criteria. Like in the case of the exploratory stage, the criteria used in these two selections are related to the grasp synthesis strategy and the robotic setup. Their specification is left open too, providing in this way a framework to the use of different strategies and robotics setups.

Figure 6 provides a filter-based outline of this stage. The overall behavior can be briefly described as follows:

- *Along the dorsal pathway.* Using the affordances selected during the exploratory stage, recovered from V3a, the AIP module looks for *compatible surface sets*. Similarly to the selected grasp surfaces, they may be ranked according to their degree of fulfillment of some given quality criteria.
- *Along the ventral pathway.* The volumetric reconstruction obtained in the exploratory stage, provided by V3 and V4, is used by LO to try to recognize the object. The output of this module is thus a model of the recognized object, an indication of the degree

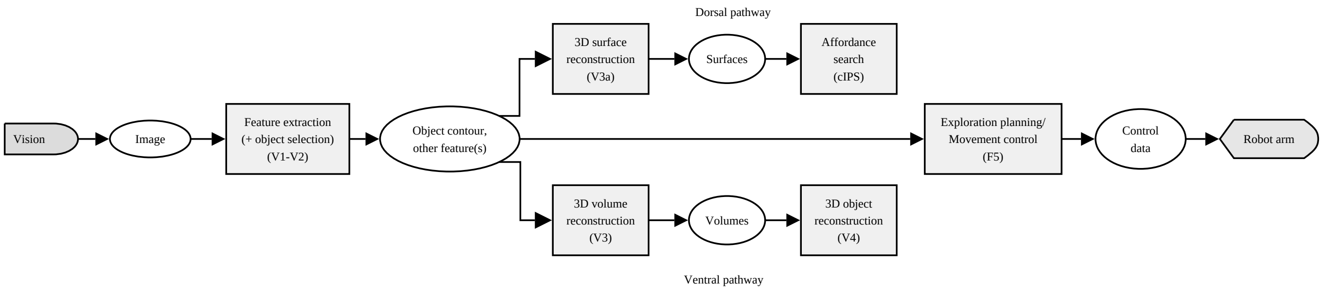

Fig. 5. Exploratory movement for the grasp synthesis (state *Exploration movement*).

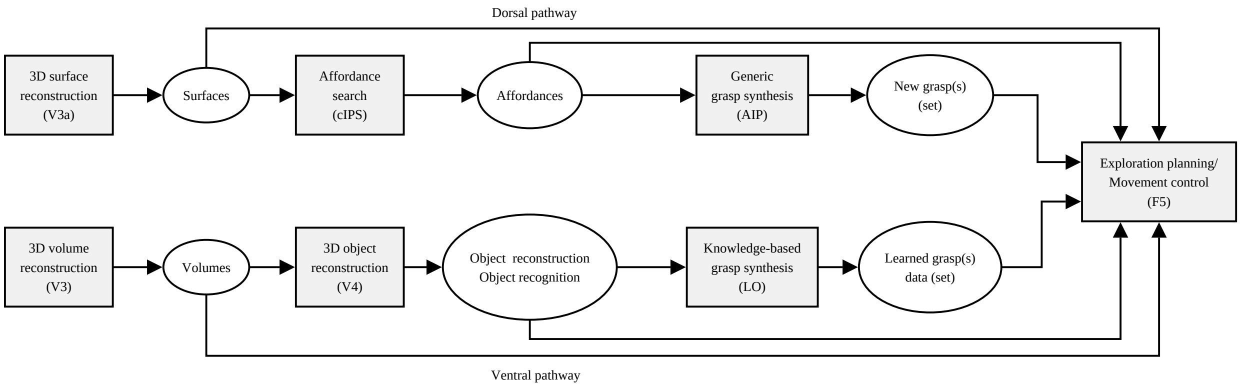

Fig. 6. Exploration planning for the grasp synthesis (state *Exploration planning*).

of confidence of the recognition and a set of grasps associated to that model.

- *Deliberation.* Algorithm 1 provides an outline of the behavior of this module (F5). Essentially, it analyzes the object information stored and produced along the dorsal and the ventral pathways. In case more information about the object is required, a new exploratory movement is planned accordingly, based on the available object data.
  If the system decides that it has enough information in order to select a grasp, it checks if it is possible to perform such a selection. If so, then the system analyzes the grasp information produced along the dorsal and ventral pathways in order to perform the grasp selection. In addition to the grasp, a corresponding motor action, out of a basic vocabulary of actions, and a grasp preshape for the hand are selected [1].

Algorithm 1 Exploration planning

```
Analyze the object information
if more information is required then
  Plan a new exploratory movement
  Trigger event Planning performed
else
  Analyze the grasp information
  if it is possible to select a grasp then
    Select a grasp
    Select a motor action and a pregrasp
    Trigger event Grasp selected
  else
    Trigger event Grasp synthesis failed
  end if
end if
```

VI. Conclusion

The goal of this paper has been to provide a framework for the development of robotic applications on the synthesis and execution of grasps. For this purpose, we have considered a biological model of grasping in humans, and a behavior-based architecture has been proposed as a tool for developing it and adapting it to the limitations of a robotic setup. Within this framework, a grasp synthesis model, tightly integrated with a vision-based 3D object reconstruction, has been introduced that can be customized to hold different synthesis strategies and robotic setups.

The proposed grasp synthesis can be extended within this framework to obtain a more detailed development of biological models. In addition, thanks to the support of task coordination, it can be integrated not only within a more complex manipulation task, but can also be coordinated with modules controlling other parts of the robot or other robots, such as in the case of the integrated control of a robot arm and a mobile platform or of two arms.

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.

ACKNOWLEDGMENT

This work has been partially funded 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). We wish to thank Mel Goodale for the enlightening discussions.

REFERENCES

[1] M. Jeannerod, M. Arbib, G. Rizzolatti, and H. Sakata, “Grasping objects: the cortical mechanisms of visuomotor transformation,” *Trends in Neuroscience*, vol. 18, no. 7, pp. 314–320, 1995.

[2] A. Milner and M. Goodale, *The visual brain in action*. Oxford University Press, 1995.

[3] T. James, G. Humphrey, J. Gati, R. Menon, and M. Goodale, “Differential effects of viewpoint on object-driven activation in dorsal and ventral streams,” *Neuron*, vol. 35, no. 4, pp. 793–801, Aug. 2002.

[4] K. Valyear, J. Culham, N. Sharif, D. Westwood, and M. Goodale, “A double dissociation between sensitivity to changes in object identity and orientation in the ventral and dorsal visual streams: A human fmri study,” *Neuropsychologia*, p. (in press), 2005.

[5] A. Singhal, E. Chinellato, J. Culham, and M. Goodale, “Dual-task interference is greater in memory-guided grasping than in visually guided grasping,” *Journal of Vision*, vol. 5, no. 8, p. 361a, 2005. [Online]. Available: http://journalofvision.org/5/8/361/

[6] 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.

[7] E. Rolls and G. Deco, *Computational Neuroscience of Vision*. Oxford, UK: Oxford University Press, 2002.

[8] 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.

[9] 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.

[10] K. Lakshminarayana, “Mechanics of form closure,” ASME, Tech. Rep. 78-DET-32, 1978.

[11] 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.

[12] 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.

[13] 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.

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

[15] C. Moore and S. Engel, “Neural response to the perception of volume in the lateral occipital complex,” *Neuron*, vol. 29, no. 1, pp. 277–286, 2001.

[16] K. Tsutsui, H. Sakata, T. Naganuma, and M. Taira, “Neural correlates for perception of 3D surface orientation from texture gradient,” *Science*, vol. 298, no. 11, pp. 409–412, Oct. 2002.

[17] 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.

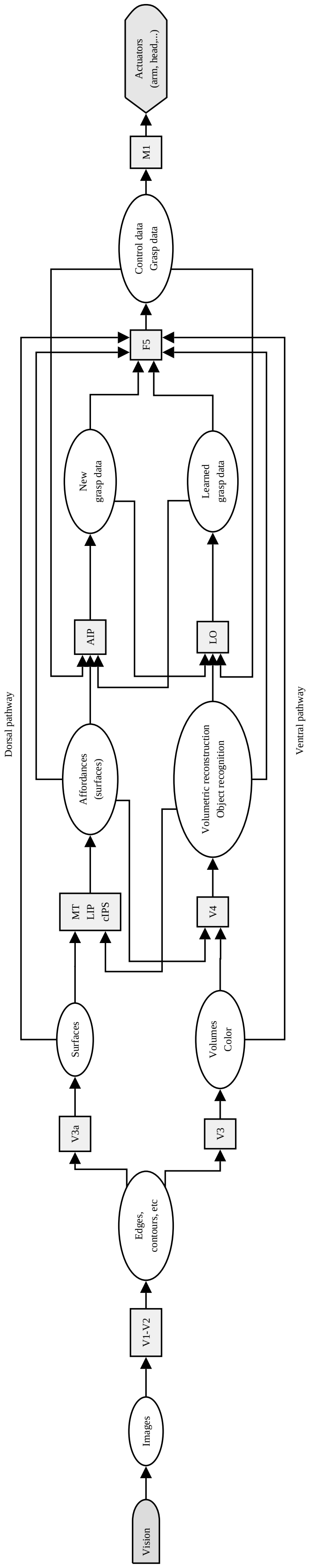

Fig. 7. Development of the grasping model from fi gure 2 using the FBA architecture.