Chapter 3
Intelligent Robotic Grasping?

Mutual interest between the fields of robotics and cognitive sciences has been steadily growing in the recent years, especially through the bridging of artificial intelligence research. Nevertheless, the differences in goals and methodology and the lack of a common language make of true interdisciplinary research still a pioneering work. As the following review will expose, grasping is not an exception to this situation.

A brief description of traditional and bio-inspired research in robotic vision-based grasping is presented and critically discussed in this chapter, with the purpose of defining a few important guidelines required to achieve fruitful cross-disciplinary research. The chapter includes a proposal for grounding symbolic representations in sensorimotor interactions for robotic grasping, synthesized in Sect. 3.3.

3.1 Vision-Based Robotic Grasping, A Brief Outline

The field of robot grasping and manipulation has been steadily growing and developing, as can be noticed comparing the most important research of the near past. The fundamental studies of the eighties (Cutkosky 1985; Mason and Salisbury 1985; Nguyen 1988) defined the basic concepts, both physical and technological, on which most of the later research built upon. In parallel, classical works on grasping in humans and primates were published (Napier 1983; Iberall 1987). A view of the robotic grasping problem more related to biological research and artificial intelligence goals and techniques was developed in the early nineties (Venkataraman and Iberall 1990; Stansfield 1991; Bekey et al. 1993). Interdisciplinary research became more common, and works such as Mackenzie and Iberall (1994), with a real interdisciplinary stance, is still among the most cited in robot grasping. Nevertheless, engineering issues and approaches clearly keeps dominating the field (Shimoga 1996), and the currently most cited reviews on the subject, Bicchi (2000) and Okamura et al. (2000), are purely technical.

The survey of Bohg et al. (2014) reviews the most recents advances on grasp planning using data-driven methods. Grasping techniques are classified depending

E. Chinellato and A.P. del Pobil, *The Visual Neuroscience of Robotic Grasping*,
Cognitive Systems Monographs 28, DOI 10.1007/978-3-319-20303-4_3

on whether the target objects are previoulsy known, unknown or similar to a familiar category. Up to six different functional architectures are identified for organizing and structuring the necessary parts of a grasping system. For what concerns the planning of grasping actions through the use of visual information, a number of simplifications have commonly been used in robotics. The most common approach is to start from a model of the object to grasp, obtained or defined in an off-line stage, and to perform object recognition if necessary. A *grasp synthesis* process (see box below) can be thus performed on the model (Lopez-Damian et al. 2005), or pre-defined grasp programs can be accessed (Taylor and Kleeman 2004). Felip et al. (2013) propose a functional organization of grasping which relies on the concept of manipulation primitive. A primitive is an independent controller focused on a specific manipulation action. Primitives are coordinated through automatas which implement complete manipulation tasks. The same plan can be used on different hardware embodiments to execute the same task.

In the robotic literature (Bicchi 2000), a *grasp* is usually defined as the set of locations (points or regions) on an object surface where the effector—a simple gripper or an artificial hand—has to contact the object for grasping it. A grasp definition may include the configuration of the hand, which depends on its kinematics and *degrees of freedom* (DOF). *Grasp synthesis* is the problem of determining a proper set of contacts on the target object, and a corresponding suitable hand configuration. See Sahbani et al. (2012) for a review of grasp synthesis techniques. The inverse problem, *grasp analysis*, involves the study and evaluation of a given grasp.

The object model can be approximated with a set of shape primitives (Miller et al. 2003), or a 3D representation useful for grasp synthesis can be built performing a visual reconstruction of the object. Due to the implicit complexity of the task (Chaumette et al. 1996), visual reconstruction is often performed using range data (Ade et al. 1995; Rutishauser and Stricker 1995). As in these works, grasp synthesis often consists of searching for antipodal regions on the reconstructed object surfaces. Otherwise, the 3D model can be decomposed into basic structural components and grasps on object parts can be either generated (Goldfeder et al. 2007) or retrieved, according to predefined preshape primitives (Miller et al. 2003). Works that try to bias visual analysis with grasping aspects, without relying on models, have often dealt with simplified worlds, such as sets of planar objects (Stanley et al. 2000; Morales et al. 2006).

Only a few ambitious works have ever been developed for visual inspection of real 3D objects aimed at grasping actions. Taylor et al. (1994) deal with the problem of how a possible grip changes with the viewpoint, whereas Cipolla and Hollinghurst (1997) look for planar surfaces within a set of possible target objects and perform a grasping action with a parallel gripper. In Seitz (1999) a visual system is introduced that moves around a target object and is able to identify its major axis and extract a symbolic representation of the contour. On such representation, suitable grasping features, both parallel and curved, are searched for. The system of Saxena et al. (2008) learns to match visual object features to suitable grasping postures, showing a high adaptability and generalization capabilities in grasping with a parallel jaw-gripper. Preliminary results of the same method in which grasping experiments are executed with a three-finger Barrett Hand are also provided, and show that the technique is promising but need adaptation to the hand kinematics. Industrial applications of autonomous vision-based grasping, such as Sanz et al. (2005), are very rare. For the cases in which a set of candidate grasps are generated, quality measures focused on aspects related to the object and effector geometry and kinematics are often used in order to assess and select the most reliable and stable grasp configurations (Markenscoff et al. 1900; Ferrari and Canny 1992; Ponce and Faverjon 1995; Xiong et al. 1999; Borst et al. 2004; Chinellato et al. 2005).

Visual control of reaching and grasping movements and of manipulation actions is a very active area, often referred to as "visual servoing". The basic principle is that the movement follows a vision-based control law, with respect to either the Cartesian space or the image space (Hutchinson et al. 1996; Cervera et al. 2003; Recatalá et al. 2008a). These methods are not focused on the visual analysis of the object, but rather on the transport action and the fine adjustment of the movement. Such detailed visual tracking of the arm and hand movements is usually not performed by primates, unless in the case of sudden changes in the environment, or for very fine manipulation tasks. The most common solution in nature is the use of ballistic movements supported by extremely reliable proprioceptive and tactile information.

In fact, the use of visual information alone, without touch feedback, is as limiting in robotics as it is in primates. Nowadays, many works that deal not only with the synthesis but also with the execution of grasping actions, make use of force or tactile sensors on the fingers of the robot hand (Hollerbach 2000; Tegin and Wikander 2005). For instance, in Platt et al. (2002), force-based controllers are concurrently used to find an appropriate placement of the fingers on the object, and in Natale and Torres-Jara (2006) a robotic hand performs tactile exploration of the object, based on a set of exploratory primitives, in order to find a good grasp. More rarely, tactile information is used to recognize an object (Allen and Roberts 1989; Petriu et al. 2004) while active haptic exploration for detecting and modeling object features is practically a novel area in robotics (Petriu et al. 1992; Johnsson and Balkenius 2006). In any case, multimodal integration is a rapidly developing topic both at the practical and at the theoretical level (Pouget et al. 2002; Prats et al. 2013), and represents a fundamental issue for the future of grasping (Coelho et al. 2001; Lippiello et al. 2006; Grzyb et al. 2008; Prats et al. 2009) and robotic research in general (Barakova and Lourens 2005; Wermter et al. 2005).

3.2 Biological Inspiration for Robot Grasping and Manipulation

Many international research projects and meetings around the world are nowadays devoted to the interplay between robotics and life sciences (Dario et al. 2005). The use of robotic solutions for medical applications is the most developed and promising direction of such interdisciplinary approach, from aided microsurgery to prosthetics. Regarding upper limb mobility and grasping, brain guidance of artificial arms and hands has been performed with monkeys (Carmena et al. 2003), and the technology is being developed in order to achieve the same for humans (Andersen et al. 2005; Acharya et al. 2007). Pursuing this goal, very complex prosthetic hands are being constructed which join the dexterity of the most advanced robotic hands with new material technology for providing them with the best comfort and aspect (Buterfass et al. 2001; Huang et al. 2006; Cipriani et al. 2008). A review of the field, completed by a biomechanical comparison between the human hand and state of the art robotic manipulators provides a valuable framework for the study of the grasp both in nature and engineering (Leon et al. 2014).

For what concerns another aspect of the interplay between robotics and neuroscience—such as the use of insights regarding brain functions and natural solutions in general in order to implement better robotic systems—biologically inspired robotics is a rapidly developing field (Bar-Cohen and Breazeal 2003; Habib et al. 2007). Nevertheless, the limited literature about biological inspiration in visual-based robot grasping at the functional level suggests that the subject has still much to offer. In fact, for most biology-inspired grasping research available in the literature, the link with neuroscience is usually a general inspiration with limited impact on the final implementation (Kragic and Christensen 2003; Laschi et al. 2006). On the opposite end are works that appear as more biologically plausible, but which relation to life sciences is not clarified. For example, Kamon et al. (1998) propose a quite flexible and adaptable learning framework for grasping, but do not explicitly cite or mention any neuroscience research or biological inspiration. Systems that do implement neural mechanisms, such as Hoffmann et al. (2005), are rare, and work only with simplified environments and ad-hoc conditions.

In the development of systems which model or imitate natural skills, either cognitive or practical, two approaches are usually cited as contrary options: bottom-up and top-down. Modern robotics favors the former, in which complex behaviors emerge from simple, hierarchically organized ones (Brooks 1999). According to the paradigm of behavior-based robotics (Arkin 1998), complex tasks can be executed through the composition of simpler ones in a bottom-up direction, to the extent that some basic components are simple stimulus-effect reflexes of the agent with the environment (Braitenberg 1984).

Behavior-based grasping and manipulation robotic research is expanding (Zöllner et al. 2005; Recatalá et al. 2008b), but the paradigm is less widespread in grasping than it is in other areas of robotics. In fact, although some kinds of problems such as basic navigation of obstacle avoidance can be solved with simple and elegant behavior-based solutions, the high complexity of the vision, motor and especially the sensorimotor processing behind grasping actions are hard to be dealt with following such techniques.

While behavior-based robotics seems to follow the right direction toward mimicking biological systems, epigenetic robotics takes one step further in approaching natural intelligence, trying to evolve high-level abilities from basic ones as in an infant developing process (Berthouze and Goldfield 2008). This developmental process can be catalyzed by imitation or human teaching (Schaal 1999; Billard and Mataric 2000; Maistros and Hayes 2000; Ito et al. 2006). For what concerns manipulation, if not proper grasping, "baby" robots which learn simple manipulation skills from exploration and observation of their actions have been already built following this trend (Metta and Fitzpatrick 2003; Natale et al. 2005), and a possible future development is imitation between robots, which could accelerate the learning process without the need of humans to intervene. Some imitation processes used for robotic systems are also biologically plausible (Demiris 2002), as they are inspired by the mirror system of the primate brain, introduced in Sect. 2.3.3, and by the concept of *motor primitives* (see box).

Motor primitives are basic motor components used to generate more complex behaviors in all kinds of movements. The composition and functionality of the motor cortex described in Sect. 2.3.3 supports their use in primate visuo-motor transformations. Motor primitives can be extracted from the thorough analysis of human motion (Drumwright et al. 2004), and used to form a motor vocabulary that can be employed to produce complex movements and action sequences (Nori and Frezza 2004).

Similarly to what happens in animals, the composition of simple behaviors as motor primitives can endow robotic systems with notable skills, dexterous manipulation being one of them (Mataric 2000). On this line, Kyota et al. (2005) use artificial neural networks to match simple voxel-based object representations to grasp configurations derived from basic human hand postures learnt with a data glove.

Even though the bottom-up approach has been usually considered more plausible from a physiological point of view, backprojections from associative to primary areas are widespread in the brain. Examples of bidirectional links, such as the premotor-parietal circuit, or the recurrent connections between visual areas, were given in the previous chapter. A mixed approach in which bottom-up mechanisms are triggered by top-down, cognitive style information, is therefore more faithful to the neurological reality. In artificial systems, whilst top-down solutions are implemented following knowledge engineering methodologies, bottom-up approaches are often coded with connectionist methods, more or less inspired by biological neural networks. Nowadays—and until artificial neural systems will resemble more closely the natural networks of neurons—more intelligent systems are probably better achieved with mixed approaches, in which pre-wired knowledge complements the connectionist mechanisms.

With the goal of linking bottom-up approaches with top-down cognitive representations, one of the most important issues is that of associating consistent symbolic meanings to aspects of the environment or of the agent-environment interaction. This issue is commonly known in cognitive science research as the *symbol grounding problem*. Though proposals have been put forth to solve the symbol grounding problem through robotic sensorimotor interactions, only little progress has been achieved with actual working systems. In the next section, a possible solution to the problem based on manipulation and grasping research is provided. Such approach can be useful in order to endow a robotic system with an implicit ability in merging practice with conceptual reasoning and thus "interpreting" its actions.

3.3 Symbol Grounding Through Robotic Manipulation

In this section, the problem of symbol grounding is addressed in the context of robotic manipulation. The goal is to obtain a more natural integration between the top-down and the bottom-up approaches by showing that there are symbols which do not refer simply to physical objects, but rather to the interactions between the robot and the objects in its environment. The description of two grasping and manipulation experiments performed at the Robotic Intelligence Lab of Universitat Jaume I, and summarized in this section, serves as a base on which to build a theory of symbolic representations for physical interactions. It will be shown that the symbols related to the interaction between agent and target object can be directly inferred from the execution of a planned action, and that neural networks can provide a suitable method for mapping complex perceptual signals to symbols. The proposal is sustained by important neuroscience studies, especially those related to the two streams of the visual cortex and the mirror system. Implementation details and further considerations can be found in Chinellato et al. (2007).

3.3.1 Symbol Grounding and Neuroscience

The symbol grounding problem is a classical challenge for cognitive science (Harnad 1990). In a traditional AI system, symbol interpretation is not intrinsic to the system, and the meaning of a symbol is always given by an external interpreter (e.g. the designer of the system). A really "intelligent" system should be able to assign symbolic meaning to an object or an action without being thought, as human beings can normally do. Harnad (1995) and del Pobil (1998) suggest that robotic sensorimotor interactions can represent a solution to this problem. In a cognitive robotic system, the symbols could be grounded in the system own capacity to interact physically with its environment.

So far, progress in symbol grounding by means of actual working robotic systems has been mostly related to visual recognition of individual objects. There is nevertheless an additional class of symbols, fundamental for cognitive robotics, which do not refer simply to physical objects but rather to the embodied physical interactions between the robot itself and the objects in the world. This kind of symbols would be more related to sensorimotor transformations, and they seem to have appeared in the evolutionary landscape long before vision as "sight" (Goodale and Westwood 2004).

The strategies employed by the brain when we interact with the world, and the more or less explicit symbolic meanings assigned to such interactions are an insightful source of inspiration for dealing with the symbol grounding problem in artificial agents. As explained above, motor primitives show different levels of complexity, and compose hierarchically to form a motor vocabulary. Complex movements and action sequences are composed in an almost linguistic way from this motor vocabulary. The question is whether motor and visuomotor primitives can be considered as symbols, extending the symbol grounding problem to a larger domain.

As explained in Chap. 2, visual processes related to specific actions in primates are different from visual processes not explicitly oriented to interaction of the subject with the environment. Area AIP codifies visual information in a grasp oriented way, associating visual features of a target object with a specific joint configuration suitable to guide the movements for grasping them, movements that are stored in the premotor cortex (Sakata et al. 2005). According to recent findings, in the inferior parietal lobule, and very likely even in AIP itself, actions are coded not only in a pragmatic way, but also in a semantic one (Gallese 2007). The activation of the dorsal stream areas of the mirror system, both parietal and premotor, when a subject is looking at an object with the purpose of interacting with it (e.g. reaching, hitting, pushing, grasping) seems to represent a "potential action". The emerging relation between sensory information and motor response may thus represent a symbolic correspondent of a grasping action (Hamilton and Grafton 2006). The dorsal stream involvement in action recognition suggests that actions are indeed associated with a symbolic meaning (Culham and Valyear 2006).

The lessons that robotics researchers can draw from these findings are: (1) the plausibility of a "dorsal style" visual elaboration that is exclusively dedicated to acting purposes; (2) the need of an emergent attribution of semantic meaning to synthesized grips. The next section introduces a robot system capable of object grasping in which visual analysis is dedicated to action, as in the primate dorsal stream, and in which a grasp codifies a relation between the hand and the object, as in humans.

3.3.2 An Emerging Categorization of Synthesized Robot Grips

Within the context of a robotic application for grasping and manipulation in a semi-structured environment, a framework was defined for characterizing candidate grips in a natural way, according to their properties in relation with the execution of a grasping action. Clustering of the candidate grip configurations occurs due to implicit

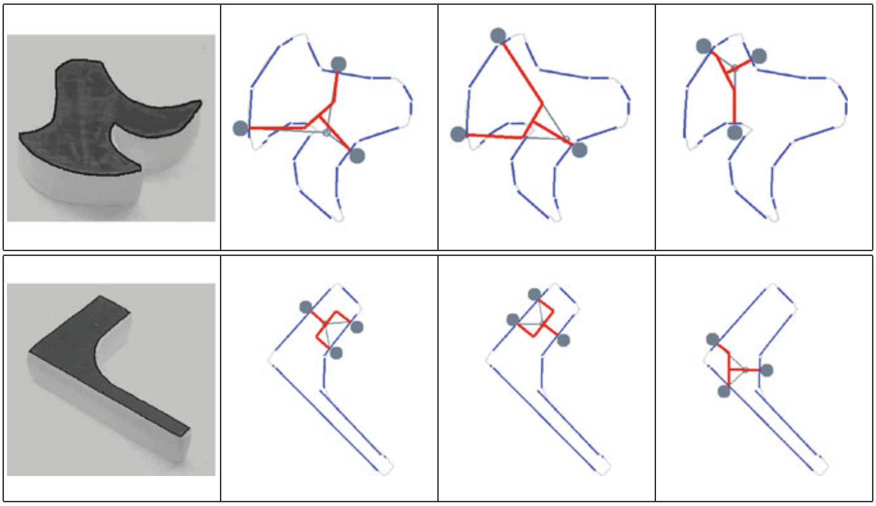

Fig. 3.1 Examples of three hand configurations found for two different objects

properties of grips, and an eventual symbolic meaning emerges through interaction of the agent with its world. The experimental setup used for the experiments described here is the one defined in Sect. 5.4.1.

The grasping action can be subdivided in four main modules or steps:

1. **Image processing:** The stereo vision system of the robot estimates the two-dimensional location of an unknown planar object placed on the table, which is the grasping target. A monocular image of the object is provided and analyzed to extract the object contour and identify on it regions suitable for finger contact.
2. **Grasp synthesis:** Several feasible candidate grasps (see some examples in Fig. 3.1) are generated, by selecting the target contact points for each triplet of grasp regions. Hand configurations for reaching the target points are computed taking into account the kinematic and geometric constraints of the Barrett Hand.
3. **Grasp evaluation:** The candidate grasps are characterized according to properties related to the geometry of the target object, the kinematics of the hand and their interaction, in order to perform a reasoned selection of the grasp to execute.
4. **Execution:** The hand is preshaped and positioned above the object, it moves down, closes the fingers so that the object is grasped, lifted and transported to a designated location. All this is performed with support of visual and tactile feedback.

Regarding grasp evaluation (step 3), a characterization scheme which includes nine high-level features was defined for providing an object-independent way to describe candidate grasps. In this way, each grasp is represented by nine measurements and thus by a point in a *9-dimensional* space. The features regard visual and motor aspects related to properties of the object contour, hand kinematics and nature of the contacts. The nine features derive from a set of quality criteria defined for assessing candidate grasp configurations (Chinellato et al. 2005).

A practical measurement of the reliability of a grasp was defined to characterize candidate configurations. After selecting a candidate grasp to execute, if the robot has been able to lift the object safely, a set of stability tests are applied in sequence. They consist of three consecutive shaking movements of the hand which are executed with an increasing acceleration. After each movement the tactile sensors are used to check whether the object has been dropped off. In this way the stability of the current grasp is measured.

An extensive experimental data gathering was realized following this protocol, and provided a qualitative measure of the success of many different grasps. A learning schema has been applied that makes use of the hyperspace described by the nine characterizing features, and tries to predict the reliability of a novel grip exploiting the information previously gathered on the outcome of already executed grasping actions. Different predicting methods have been applied and compared, with quite satisfying performances (Chinellato et al. 2003; Morales et al. 2004). These results showed that grasp reliability could be implicitly related to the characterizing features.

Summarizing, in this example each grip is represented as a point in a multidimensional space, and a procedure is introduced to predict a query point based on its similarity to previous grasping experiences. This is a case of *instance-based*, also known as *memory-based* learning (Aha 1997), which is a numeric version of the explicitly symbolic *case-based reasoning* (Waltz 1995). Although in this approach there is no explicit representation of the target function when training samples are provided, a natural characterization of the grips evolve from experience. The "character" of a new grip is recognized from the system as similar to some experience it already passed through, and the validity of such emerging associating ability is proved by the consequent predicting capacity of the system.

The system thus assigns each novel grip a symbolic value originating directly from its visuomotor character. Also, the visual analysis performed is only oriented toward action related features, as in the cortical dorsal stream, and there is no recognition of the target object. In this way, symbols are not related to the identity of objects, but rather to the different sensory experiences that objects can provide.

The next section proposes a procedure for extracting symbolic meaning from sensorimotor interactions through the use of connectionist methods.

3.3.3 Extracting Symbolic Meanings from Physical Interactions

In the previous section, a symbolic representation of sensorimotor experience was used to compare grips and predict the reliability of upcoming actions. This section describes an approach for extracting explicit symbolic meaning from sensor data

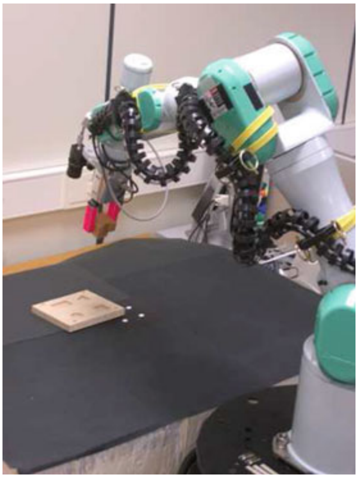

Fig. 3.2 System setup for the peg-in-hole task

through the use of artificial neural networks. In this case, the symbolic relation refers to the contact states between an already grasped object and the goal position.

The study is based on the classical two-dimensional peg-in-hole insertion task (Cervera and del Pobil 2000), performed using the setup depicted in Fig. 3.2. The task is to insert a peg hold with the gripper in a chamferless hole. Wrist-mounted force sensors are used to measure the external forces and torques applied to the peg as the manipulator interacts with the environment.

Symbol grounding in this experiment is required to assign symbolic values to the physical interactions between the robot—the peg can be considered as a prolongation of the robot gripper—and the environment. The correct identification of such interactions is of fundamental importance for the adequate execution of the task: in this particular case the insertion of the peg into the hole. Physical interactions are modeled as contact states between peg and hole, that have to be identified with the help of the force sensors. In Fig. 3.3, all possible contact states and the corresponding forces are shown. The no-contact state, **F0**, shows the weight force, while all the others only show the reaction forces, but weight is considered too. In order to identify the contact states, only the direction of forces is relevant, not their magnitude. The sensors provide only three raw signals of force and torque, and the problem is to

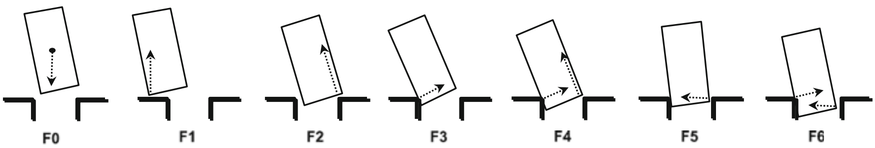

Fig. 3.3 Set of contact states for the peg-in-hole task, showing the contact forces between peg and hole. State **F0** shows only the weight force which is omitted in all other graphs

Table 3.1 Contact states classification percentages

State	Right	Wrong	Unknown
F1	94	5	1
F2	100	–	–
F3	59	34	7
F4	97	–	3
F5	100	–	–
F6	78	21	1
Average	88	10	2

appropriately map these signals to one symbolic contact state. This is not a trivial problem due to the variability of the forces and the superposition of the peg weight and one or several reaction forces.

In order to perform the matching between sensory data and symbolic states, Self-Organizing Maps were used (Kohonen 1990). The experiments consist of three phases:

1. **Training.** The neural network is trained with a set of 1200 random input samples equally distributed across contact states, according to the procedure described in Kohonen (1990). In this way the net learns in an unsupervised way the natural structure of the data.
2. **Calibration.** A set of 600 labeled samples is used for calibrating the net to the contact states. The network response is analyzed for all inputs, and each network unit is labeled with the contact state for which it has been more frequently selected as the closest unit.
3. **Testing.** The performance of the network is tested with an independent set of 600 samples. For each sample, the most responsive unit is selected, and the output contact state is given by that unit's label. An uncertain response occurs if the unit is unlabeled.

Results from a typical experiment are shown in Table 3.1. Two states, **F2** and **F5**, are perfectly identified. State **F4** is correctly identified in the 97 % of cases and **F1** is properly classified in the 94 % of the cases. States **F3** and **F6** seem to be more difficult to identify, and the correct classification percentages are smaller. The average network performance, 88 % success, is anyway very good, and it has to be taken into account that only force information has been used.

After establishing the correspondence between force sensor data and symbols denoting contacts, the peg-in-hole insertion problem can be solved in a number of ways. These are depicted in Fig. 3.4, that shows a perception-based diagram in which the increments inside the nodes denote the actions to be performed for reaching the

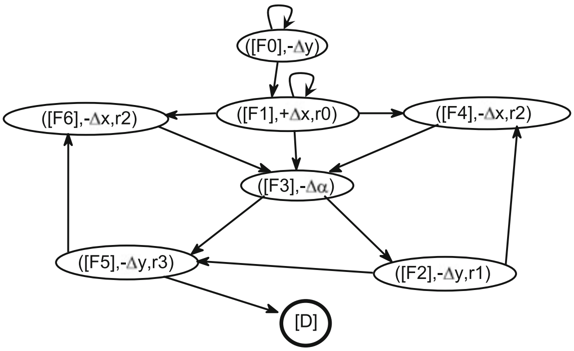

Fig. 3.4 Perception-based motion plan for solving the peg-in-hole task. The increments within the nodes symbolize the possible actions and the arrows the consequent state transitions

neighbor nodes. The transitions are defined so that the goal position **D** can be reached from any state. Several tests were performed with a real robot and sensors using the transition rules of Fig. 3.4, and following different sequences of contact states. The good experimental results demonstrated that the task could actually be performed by using direct sensory data only, without the need for information about the position and orientation of the peg.

This example shows how self-organizing neural networks were able to automatically extract a number of symbolic states that represent relevant real-world situations useful in a transition graph for the solution of a relatively complex problem. Most importantly, such symbolic knowledge derived directly from sensory information regarding ongoing actions.

3.3.4 Symbolic Value of Hand-Object Interactions

The two case studies exposed above show how symbolic meanings can naturally arise from the physical interaction between an agent and objects in its environment, suggesting that motor primitives constitute a consistent source of symbolic knowledge.

Looking once more at cut-edge research in neurophysiology, the existence of a parieto-premotor mirror system (Sect. 2.5) supports the idea of extending the symbol concept to motor behaviors and sensorimotor interactions. In fact, the symbols managed by AIP do not codify objects, but action-oriented visual representations, and association patterns between objects and distal subject effectors (Tunik et al. 2007). Taking a step further in the analysis of the mirror mechanisms, it has been proposed that symbolic communication and language evolved from a neural motor system involved in action recognition (Keysers et al. 2003; Hamilton and Grafton 2006). Moreover, the mirror system seems to play a critical role in learning by imitation, a skill that we are only beginning to develop in robots. Broca's area in humans, traditionally related to language production, is the most likely correspondent of F5, where mirror neurons were first discovered (Binkofski and Buccino 2004; Grèzes et al. 2003). This would confirm that action understanding, recognition and mental imagery of actions do not differ conceptually from object recognition or imaging, and that complex cognitive processes emerge from simple behaviors which firstly evolved in order to endow the organism with skills for better interacting in its environment.

3.4 Toward Intelligent Robotic Grasping

In the previous section, the existence of symbols that are grounded in the physical sensorimotor interactions between the robot and the objects in the world has been discussed. The two presented case studies show how symbolic meaning can describe, and be extracted from, visuomotor experiences regarding manipulation tasks, allowing the robot to build a representation of the possible interactions with its surrounding environment. The merging of different sensory modalities, such as vision and touch, in the exploration of the environment is probably the most necessary further development of the proposed approach. A fundamental conclusion, consistent with findings about the action-oriented human dorsal visuomotor stream, is that in a robotic system there is no need to model, recognize or classify an object in order to physically interact with it, since the symbols derived from sensorimotor experience identify particular physical interactions between the robot hand and the target object.

As previously explained, classical robotic approaches do not follow this path. Instead, they consider either object recognition or full visual reconstruction previous to grasp analysis, which is normally based on an object model (Bicchi 2000). This pattern corresponds to the scheme of Fig. 3.5a, in which perception-synthesis-action

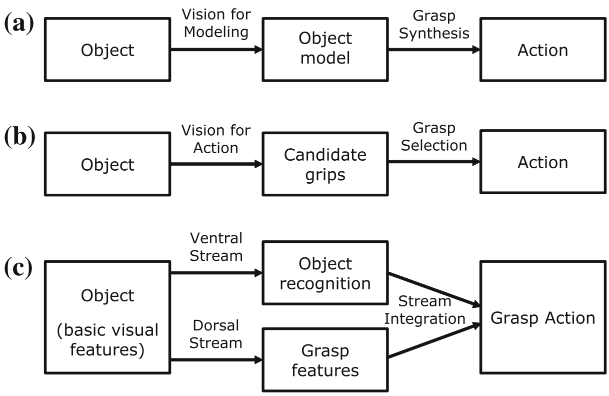

Fig. 3.5 Different approaches in vision-based grasping: **a** is the traditional perception-reason-act paradigm; **b** represents an action-based vision approach; **c** is the integrated methodology proposed in this work

are sequentially executed, and vision is general purpose rather than goal-oriented. There are works in the literature which exploit visual features in a manner focused on grasp purposes, but without the aid of "cognitive" information about objects (see e.g. Morales et al. 2006; Saxena et al. 2008 and Sect. 3.3.2). This approach is symbolized in Fig. 3.5b. The first method lacks of flexibility, as it builds on a nearly general-purpose visual processing stage. The second method partially resembles the job of the dorsal stream, and is thus more biologically plausible and often very efficient. Nevertheless, it does not exploit the kind of cognitive information that makes human grasping skills largely transcend a geometric or kinematic analysis.

The review presented in this chapter shows that the grasp problem in robotics is still largely unsolved. The proposal put forth in this book for obtaining a more reliable and meaningful interaction with real world objects is to integrate in a robotic grasping system instantaneous visual information gathered in an action oriented-manner (dorsal pathway) with experience-mediated information (ventral pathway).

In Fig. 3.5c a representation of this solution, which somehow encompasses the two previous ones, is depicted. This integration process is what human beings do all the time: we join our knowledge about the objects we are going to grasp, and the experience of previous actions, with the analysis of the actual, concrete situation we are facing. The model introduced in the next chapter goes toward this direction, and analyzes the indications of neuroscience research from a practical point of view, in order to devise a set of mechanisms able to endow a robot grasping system with increased capabilities in interacting with its nearby environment.

References

Acharya S, Aggarwal V, Tenore F, Shin HC, Etienne-Cummings R, Schieber M, Thakor N (2007) Towards a brain-computer interface for dexterous control of a multi-fingered prosthetic hand. In: International IEEE/EMBS conference on neural engineering, pp 200–203. doi:10.1109/CNE.2007.369646

Ade F, Rutishauser M, Trobina M (1995) Grasping unknown objects. In: Bunke H, (ed) Dagstuhl proceedings seminar. World Scientific, pp 445–459

Aha D (1997) Lazy learning. Artif Intell Rev 11:7–10

Allen P, Roberts K (1989) Haptic object recognition using a multi-fingered dextrous hand. In: IEEE international conference on robotics and automation, pp 342–347. doi:10.1109/ROBOT.1989.100011

Andersen R, Andersen R, Musallam S, Burdick J, Cham J (2005) Cognitive based neural prosthetics. In: IEEE international conference on robotics and automation, pp 1908–1913

Arkin R (1998) Behavior-based robotics. MIT Press, Cambridge

Barakova EI, Lourens T (2005) Event based self-supervised temporal integration for multimodal sensor data. J Integr Neurosci 4(2):265–282. doi:10.1142/S021963520500077X

Bar-Cohen Y, Breazeal C (2003) Biologically inspired intelligent robots. SPIE Press, San Diego

Bekey G, Liu H, Tomovic R, Karplus W (1993) Knowledge-based control of grasping in robot hands using heuristics from human motor skills. IEEE Trans Robot Autom 9(6):709–722

Berthouze L, Goldfield EC (2008) Assembly, tuning, and transfer of action systems in infants and robots. Infant Child Dev 17(1):25–42. doi:10.1002/icd.542

Bicchi A (2000) Hand for dexterous manipulation and robust grasping: a difficult road towards simplicity. IEEE Trans Robot Autom 16(6):652–662

Billard A, Mataric M (2000) Learning motor skills by imitation: a biologically inspired robotic model. In: International conference on cognitive and neural systems, Boston

Binkofski F, Buccino G (2004) Motor functions of the Broca's region. Brain Lang 89(2):362–369. doi:10.1016/S0093-934X(03)00358-4

Bohg J, Morales A, Asfour T, Kragic D (2014) Data-driven grasps synthesis—a survey. IEEE Trans Robot 30(2):289–309

Borst C, Fischer M, Hirzinger G (2004) Grasp planning: how to choose a suitable task space. In: IEEE international conference on robotics and automation, New Orleans, USA, pp 319–325

Braitenberg V (1984) Vehicles: experiments in synthetic psychology. MIT Press, Cambridge

Brooks R (1999) Cambrian intelligence: the early history of the new AI. MIT Press, Cambridge

Buterfass J, Grebenstain M, HLiu, Hirzinger G (2001) DLR-Hand II: next generation of a dextrous robot hand. In: IEEE international conference on robotics and automation, Seoul, Korea

Carmena JM, Lebedev MA, Crist RE, O'Doherty JE, Santucci DM, Dimitrov DF, Patil PG, Henriquez CS, Nicolelis MAL (2003) Learning to control a brain-machine interface for reaching and grasping by primates. PLoS Biol 1(2):e42. doi:10.1371/journal.pbio.0000042

Cervera E, Pobil APD, Berry F, Martinet P (2003) Improving image-based visual servoing with three-dimensional features. Int J Rob Res 22(10–11):821–840

Cervera E, del Pobil A (2000) A qualitative-connectionist approach to robotic spatial planning. Spat Cognit Comput 1(2):51–76

Chaumette F, Boukir S, Bouthemy P, Juvin D (1996) Structure from controlled motion. IEEE Trans Pattern Anal Mach Intell 18(5):492–504

Chinellato E, Morales A, Fisher RB, del Pobil AP (2005) Visual quality measures for characterizing planar robot grasps. IEEE Trans Syst Man Cybern Part C 35(1):30–41

Chinellato E, Morales A, Cervera E, del Pobil AP (2007) Symbol grounding through robotic manipulation in cognitive systems. Robot Auton Syst 55(12):851–859

Chinellato E, Morales A, Sanz PJ, del Pobil AP (2003) Validation of features for characterizing robot grasps. In: Mira J, Alvarez J (eds) Artificial neural nets problem solving methods, LNCS 2687. Springer, pp 193–200

Cipolla R, Hollinghurst N (1997) Visually guided grasping in unstructured environments. Robot Auton Syst 19:337–346

Cipriani C, Zaccone F, Micera S, Carrozza MM (2008) On the shared control of an EMG-controlled prosthetic hand: analysis of user-prosthesis interaction. IEEE Trans Robot Autom 24(1):170–184. doi:10.1109/TRO.2007.910708

Coelho J Jr, Pieter J, Grupen R (2001) Developing haptic and visual perceptual categories for reaching and grasping with a humanoid robot. Robot Auton Syst 37:195–218

Culham JC, Valyear KF (2006) Human parietal cortex in action. Curr Opin Neurobiol 16(2):205–212. doi:10.1016/j.conb.2006.03.005

Cutkosky M (1985) Robotic grasping and fine manipulation. Kluwer Academic Press, Dordrecht

Dario P, Carrozza M, Guglielmelli E, Laschi C, Menciassi A, Micera S, Vecchi F (2005) Robotics as a future and emerging technology: biomimetics, cybernetics, and neuro-robotics in European projects. IEEE Robot Autom Mag 12(2):29–45. doi:10.1109/MRA.2005.1458320

del Pobil AP (1998) The grand challenge is called: robotic intelligence. In: Tasks and methods in applied artificial intelligence, LNCS1416. Springer, pp 15–24

Demiris Y (2002) Mirror neurons, imitation, and the learning of movement sequences. In: International conference on neural information processing, Singapore, pp 111–115

Drumwright E, Jenkins O, Mataric M (2004) Exemplar-based primitives for humanoid movement classification and control. In: IEEE international conference on robotics and automation, New Orleans, USA

Felip J, Laaksonen J, Morales A, Kyrki V (2013) Manipulation primitives: a paradigm for abstraction and execution of grasping and manipulation tasks. Robot Auton Syst 61(3):283–296

Ferrari C, Canny J (1992) Planning optimal grasps. In: IEEE international conference on robotics and automation, Nice, France, pp 2290–2295
Gallese V (2007) The "conscious" dorsal stream: embodied simulation and its role in space and action conscious awareness. Psyche 13(1):
Goldfeder C, Allen P, Pelossof R, Lackner C (2007) Grasp planning via decomposition trees. In: IEEE international conference on robotics and automation, Rome, Italy
Goodale MA, Westwood DA (2004) An evolving view of duplex vision: separate but interacting cortical pathways for perception and action. Curr Opin Neurobiol 14(2):203–211. doi:10.1016/j.conb.2004.03.002
Grèzes J, Armony JL, Rowe J, Passingham RE (2003) Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. Neuroimage 18(4):928–937
Grzyb BJ, Chinellato E, Morales A, del Pobil AP (2008) Robust grasping of 3D objects with stereo vision and tactile feedback. In: IEEE international conference on intelligent robots and systems
Habib MK, Watanabe K, Izumi K (2007) Biomimetics robots from bio-inspiration to implementation. In: Annual conference of the ieee industrial electronics society, pp 143–148. doi:10.1109/IECON.2007.4460382
Hamilton AF, Grafton ST (2006) Goal representation in human anterior intraparietal sulcus. J Neurosci 26(4):1133–1137. doi:10.1523/JNEUROSCI.4551-05.2006
Harnad S (1995) Grounding symbolic capacity in robotic capacity. In: Brooks R (ed) Steels. The artificial life route to artificial intelligence, building situated embodied agents, Lawrence Erlbaum, pp 277–286
Harnad S (1990) The symbol grounding problem. Phys D 42:335–346. doi:10.1016/0167-2789(90)90087-6
Hoffmann H, Schenck W, Möller R (2005) Learning visuomotor transformations for gaze-control and grasping. Biol Cybern 93(2):119–130. doi:10.1007/s00422-005-0575-x
Hollerbach JM (2000) Some current issues in haptics reseach. In: IEEE international conference on robotics and automation, San Francisco, USA, pp 757–762
Huang H, Li J, Liu Y, Hou L, Cai H, Liu H (2006) The mechanical design and experiments of HIT/DLR prosthetic hand. In: IEEE international conference on robotics and biomimetics, pp 896–901. doi:10.1109/ROBIO.2006.340339
Hutchinson S, Hager G, Corke P (1996) A tutorial on visual servo control. IEEE Trans Robot Autom 12(5):651–670
Iberall T (1987) The nature of human prehension: three dextrous hands in one. In: IEEE international conference on robotics and automation, pp 396–401
Ito M, Noda K, Hoshino Y, Tani J (2006) Dynamic and interactive generation of object handling behaviors by a small humanoid robot using a dynamic neural network model. Neural Netw 19(3):323–337. doi:10.1016/j.neunet.2006.02.007
Johnsson M, Balkenius C (2006) Experiments with artificial haptic perception in a robotic hand. J Intell Fuzzy Syst 17(4):377–385
Kamon I, Kamon I, Flash T, Edelman S (1998) Learning visually guided grasping: a test case in sensorimotor learning. IEEE Trans Syst Man Cybern Part A 28(3):266–276. doi:10.1109/3468.668958
Keysers C, Kohler E, Umiltà MA, Nanetti L, Fogassi L, Gallese V (2003) Audiovisual mirror neurons and action recognition. Exp Brain Res 153(4):628–636. doi:10.1007/s00221-003-1603-5
Kohonen T (1990) The self-organizing map. In: Proceedings of IEEE, vol 78, no 9, pp 1464–1480
Kragic D, Christensen HI (2003) Biologically motivated visual servoing and grasping for real world tasks. In: IEEE international conference on intelligent robots and systems, Las Vegas, USA
Kyota F, Watabe T, Saito S, Nakajima M (2005) Detection and evaluation of grasping positions for autonomous agents. In: International Workshop on Language Understanding and Agents for Real World Interaction, pp 453–460

Laschi C, Asuni G, Teti G, Carrozza M, Dario P, Guglielmelli E, Johansson R (2006) A bio-inspired neural sensory-motor coordination scheme for robot reaching and preshaping. In: IEEE international conference on biomedical robotics and biomechatronics, pp 531–536

Leon B, Morales A, Sancho-Bru J (2014) From robot to human grasping simulation, vol 19. Springer, New York

Lippiello V, Siciliano B, Villani L (2006) Robot interaction control using force and vision. In: IEEE international conference on intelligent robots and systems, pp 1470–1475. doi:10.1109/IROS.2006.281974

Lopez-Damian E, Sidobore D, Alami R (2005) Grasp planning for non-convex objects. In: International symposium on robotics, Tokyo, Japan

Mackenzie C, Iberall T (1994) The grasping hand. North Holland, New York

Maistros G, Hayes G (2000) An imitation mechanism inspired from neurophysiology. In: International workshop on current computational architectures integrating neural networks and neuroscience, Durham, UK

Markenscoff X, Li L, Papadimitriou C (1990) The geometry of grasping. Int J Robot Res 9(1):61–74

Mason M, Salisbury J Jr (1985) Robot hands and the mechanics of manipulation. MIT Press, Cambridge

Mataric MJ (2000) Getting humanoids to move and imitate. IEEE Intell Syst 15(4):18–24

Metta G, Fitzpatrick P (2003) Early integration of vision and manipulation. Adapt Behav 11(2):109–128

Miller AT, Knopp S, Christensen HI, Allen PK (2003) Automatic grasp planning using shape primitives. In: IEEE international conference on robotics and automation, Taipei, Taiwan

Morales A, Chinellato E, Fagg AH, del Pobil AP (2004) Using experience for assessing grasp reliability. Int J Humanoid Rob 1(4):671–691

Morales A, Sanz PJ, del Pobil AP, Fagg AH (2006) Vision-based three-finger grasp synthesis constrained by hand geometry. Robot Auton Syst 54(6):496–512

Napier J (1983) Hands. Princeton University Press, Princeton

Natale L, Metta G, Sandini G (2005) A developmental approach to grasping. In: Developmental robotics AAAI spring symposium

Natale L, Torres-Jara E (2006) A sensitive approach to grasping. In: International conference on epigenetic robotics, Paris, France

Nguyen VD (1988) Constructing force-closure grasps. Int J Robot Res 7(3):3–16

Nori F, Frezza R (2004) Biologically inspired control of a kinematic chain using the superposition of motion primitives. In: IEEE conference on decision and control, vol 1, pp 1075–1080

Okamura A, Smaby N, Cutkosky M (2000) An overview of dexterous manipulation. In: IEEE international conference on robotics and automation, San Francisco, CA, USA, pp 255–260

Petriu EM, McMath WS, Yeung SSK, Trif N (1992) Active tactile perception of object surface geometric profiles. IEEE Trans Instrum Measur 41(1):87–92

Petriu EM, Yeung SKS, Das SR, Cretu AM, Spoelder HJW (2004) Robotic tactile recognition of pseudorandom encoded objects. IEEE Trans Instrum Meas 53(5):1425–1432. doi:10.1109/TIM.2004.834100

Platt R, Fagg A, Grupen R (2002) Nullspace composition of control laws for grasping. In: IEEE international conference on intelligent robots and systems, Lausanne, Switzerland, pp 1717–1723

Ponce J, Faverjon B (1995) On computing three-finger force-closure grasps of polygonal objects. IEEE Trans Robot Autom 11(6):868–881

Pouget A, Deneve S, Duhamel JR (2002) A computational perspective on the neural basis of multisensory spatial representations. Nat Rev Neurosci 3(9):741–747. doi:10.1038/nrn914

Prats M, del Pobil AP, Sanz PJ (2013) Robot physical Interaction through the combination of Vision, Tactile and Force Feedback, vol 84. Springer, New York

Prats M, Sanz PJ, del Pobil A (2009) Vision-tactile-force integration and robot physical interaction. In: IEEE International Conference on Robotics and Automation, 2009, ICRA '09, pp 3975–3980. doi:10.1109/ROBOT.2009.5152515

Recatalá G, Carloni R, Melchiorri C, Sanz P, Cervera E, del Pobil A (2008a) Vision-based grasp tracking for planar objects. IEEE Trans Syst Man Cybern Part C Appl Rev 38(6):844–849
Recatalá G, Chinellato E, del Pobil AP, Mezouar Y, Martinet P (2008b) Biologically-inspired 3D grasp synthesis based on visual exploration. Auton Robots 25(1):59–70. doi:10.1007/s10514-008-9086-7
Rutishauser M, Stricker M (1995) Searching for grasping opportunities on unmodeled 3D objects. In: British machine vision conference, vol 1, pp 277–286
Sahbani A, El-Khoury S, Bidaud P (2012) An overview of 3d object grasp synthesis algorithms. Robot Auton Syst 60(3):326–336. doi:10.1016/j.robot.2011.07.016
Sakata H, Tsutsui KI, Taira M (2005) Toward an understanding of the neural processing for 3D shape perception. Neuropsychologia 43(2):151–161. doi:10.1016/j.neuropsychologia.2004.11.003
Sanz PJ, Requena A, Iñesta JM, Del Pobil AP (2005) Grasping the not-so-obvious: vision-based object handling for industrial applications. IEEE Robot Autom Mag 12(3):44–52. doi:10.1109/MRA.2005.1511868
Saxena A, Driemeyer J, Ng AY (2008) Robotic grasping of novel objects using vision. Int J Robot Res 27(2):157–173. doi:10.1177/0278364907087172
Schaal S (1999) Is imitation learning the route to humanoid robots? Trends Cogn Sci 3(6):233–242
Seitz M (1999) Towards autonomous robotic servicing: using an integrated arm-eye system for manipulating unknown objects. Robot Auton Syst 26(1):23–42
Shimoga K (1996) Robot grasp synthesis algorithms: a survey. Int J Robot Res 15(3):230–266
Stanley K, Wu J, Gruver W (2000) Implementation of vision-based planar grasp planning. IEEE Trans Syst Man Cybern 30(4):517–524
Stansfield S (1991) Robotic grasping of unknown objects: a knowledge-based approach. Int J Robot Res 10(4):314–326
Taylor M, Blake A, Cox A (1994) Visually guided grasping in 3D. In: IEEE international conference on robotics and automation, San Diego, California, pp 761–766
Taylor G, Kleeman L (2004) Integration of robust visual perception and control for a domestic humanoid robot. In: IEEE International conference on intelligent robots and systems, pp 1010–1015. doi:10.1109/IROS.2004.1389485
Tegin J, Wikander J (2005) Tactile sensing in intelligent robotic manipulation—a review. Ind Robo Int J 32:64–70. doi:10.1108/01439910510573318
Tunik E, Rice NJ, Hamilton A, Grafton ST (2007) Beyond grasping: representation of action in human anterior intraparietal sulcus. Neuroimage 36(Suppl 2):T77–T86. doi:10.1016/j.neuroimage.2007.03.026
Venkataraman S, Iberall T (eds) (1990) Dextrous robot hands. Springer, New York
Waltz DL (1995) Memory-based reasoning. In: Arbib MA (ed) The handbook of brain theory and neural networks, MIT Press, pp 661–662
Wermter S, Weber C, Elshaw M (2005) Associative neural models for biomimetic multi-modal learning in a mirror neuron-based robot. In: Cangelosi A, Bugmann G, Borisyuk R (eds) Modeling language cognition and action. World Scientific, pp 31–46
Xiong CH, Li YF, Ding H (1999) On the dynamic stability of grasping. Int J Robot Res 18(9):951–958
Zöllner R, Pardowitz M, Knoop S, Dillmann R (2005) Towards cognitive robots: building hierarchical task representations of manipulators from human demostration. In: IEEE international conference on robotics and automation, Barcelona, Spain, pp 1547–1552