

From sensorimotor development to object perception

Lorenzo Natale, Francesco Orabona, Fabio Berton, Giorgio Metta, Giulio Sandini

LIRA-Lab, DIST

University of Genoa

Viale Causa 13, 16145, Genoa, Italy

Email: {nat, bremen, fberton, pasa, sandini}@liralab.it

Abstract—This paper describes a developmental sequence that allows a humanoid robot to learn about the shape of its body and successively about certain parts of the environment. We equipped the humanoid robot with an initial set of motor and perceptual competencies ranging from simple stereotyped actions to more sophisticated visual routines providing a bottom-up attention system. This initial form of sensorimotor coordination is sufficient to initiate the interaction with the environment and allows the robot to improve its motor and perceptual skills by first constructing a “body-schema” and later by learning about objects. The body-schema allows controlling movements to fixate, reach and touch objects in the environment. The interaction is further used to form a visual model of the objects grasped by the robot which eventually modulate the attention system in a top-down way. In another experiment we show an initial effort to study the acquisition of object affordances. We discuss the importance of sensorimotor coordination as a required step not only for the control of action but also, and more importantly, for perceptual development.

I. INTRODUCTION

Manipulation is a unique opportunity to study the interaction between an artificial system and the environment. We focus on manipulation not only as a means to perform useful practical tasks but, also and in particular, because it offers the possibility to investigate active learning. Active learning refers, for example, to the ability of an agent to autonomously perform and guide the exploration of the environment. At a deeper level, action can contribute to change the environment in a direction that is best suited to the agent’s goal, for example, to facilitate perception. In this context manipulation allows the agent to collect information about objects by performing specific actions on them [1]. Even very simple forms of action like poking or pushing an object can be sufficient to this purpose [2].

The sensorial experience of a humanoid robot can be quite rich including, for example, the possibility of multi-sensory perception: some features of objects are more naturally perceived through senses other than vision. For example the smoothness of a surface, the weight of an object and its three dimensional structure are naturally determined through tactile experience. This information is extracted by appropriate exploratory actions. The importance of motor activity for perceptual development has been emphasized in developmental psychology [3], [4]. Many researchers agree on the fact that motor development during infancy determines the timing of perceptual development (for a review see [5]). For example perception of object features such as volume, hardness, texture

and weight are unlikely to emerge before 6/9 months of age. Haptic sensitivity of three dimensional shapes appears even later at around 12/15 months. It is perhaps illuminating that this timetable fits surprisingly well with the development of actions in infants: the ability to move the hand is required for infants to begin manipulating objects and consequently perceiving certain properties. Lederman and Klatzky [6] have shown that adults make use of stereotyped hand movement (*exploratory procedures*) to determine certain properties of objects; different procedures were employed by subjects to assess different properties. The ability of infants to correctly execute these procedures can determine their ability to perceive the associated object characteristic. Infants ability to interact with objects is indeed quite limited at birth; early reaching in newborns is pretty inaccurate and only rarely result in actual contact with the object [7]. At the age of three months infants are more reliable in grasping objects, although grasp is usually with the full hand open (as in the power grasp). Only later on at 6/9 months of age infants become skilled in handling objects and grasping them with differentiated grasp types [8]. Accordingly, perception of properties like temperature, size and hardness can occur relatively early in development, whereas properties requiring more dexterous actions like texture or three dimensional shape would emerge only later on.

Similarly, we are pursuing a developmental approach for the design of a humanoid robot. Development of the robot unfolds along three phases: learning about the body, learning to interact with the environment, and learning to interpret events. In the first phase the robot learns properties of its body, which allow recognizing and controlling movement. For example, the robot controls reaching movements by first learning the weight of its arm and to recognize the hand. These abilities are used in the next phase to initiate the interaction with the environment and to learn about it. The robot begins this exploration by reaching for objects and learning their physical properties when it manages to grasp them [9]–[11]. The robot’s experience acquired during this interaction is used in the third phase to interpret events around the robot by matching expectations and perceptions (as for example in [1]). We focus here on the two first phases: learning about the body and learning to interact with the environment. We describe how the robot builds an internal model of its hand which allows localizing it in the visual scene. The hand internal model is used to direct gaze towards the hand and to learn an inverse

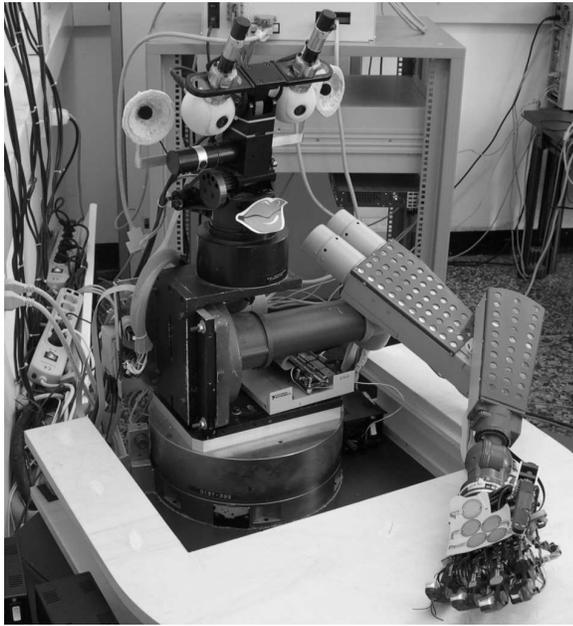


Fig. 1. The robotic platform: The Babybot.

model which can be used to control how to reach a point in space. The robot uses these abilities to build a visual model of the objects it happens to grasp. The robot's ability to interact with the environment influences the visual attention system; the visual model of the object grasped by the robot, in fact, is then used as a top-down primer during the search of graspable objects.

The rest of the paper is organized as follow. Section II describes the robotic platform. Section III describes the attention system of the robot, the used model of objects and the method to extract three-dimensional information about objects. Section IV describes how the robot acquires its motor skills. Section V presents an experiment where we show how the modules described in the paper are integrated in a complete behavioral system. Finally in Section VI we draw the conclusions.

II. THE ROBOTIC PLATFORM

The experiments reported in this paper were performed by using an upper torso humanoid robot called Babybot (Figure 1). The Babybot consists of a head, an arm and a hand. The head has five degrees of freedom, and it is equipped with two cameras, two microphones and a set of gyroscopes. The cameras can pan independently and tilt around a common axis; the remaining degrees of freedom allows the head to pan and tilt at the level of the neck. The arm is an industrial PUMA 260 manipulator. The hand is mounted on the arm end point. It consists of a total of 16 degrees of freedom actuated by only 6 motors. Its five fingers are thus largely underactuated: the thumb and index are controlled independently by two motors each, whereas the remaining two motors are connected to the middle, ring and small finger which form a single virtual joint. The coupling between each joint and the motors is achieved by means of springs which give the hand a certain degree

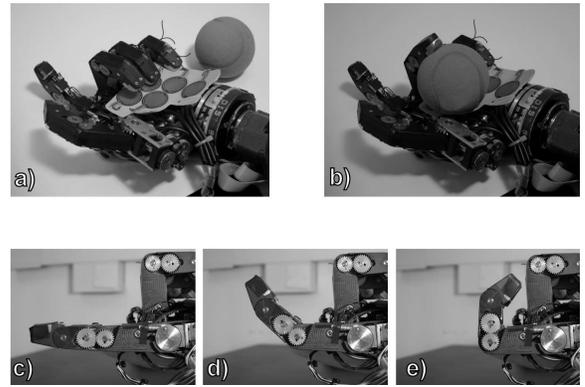


Fig. 2. Details of the hand of the Babybot.

of compliance and elasticity. Magnetic potentiometers provide position and force feedback at each joint whereas force sensing resistors on the palm and fingers provide tactile feedback (see figure 1 and figure 2). A more detailed description of the hand can be found in [12].

III. VISUAL SYSTEM

The robot visual system employs log-polar images as in [13]. The log-polar transformation, applied in this case to the traditional rectangular images coming from the cameras, mimics the distribution of the photoreceptors in the retina and the topological mapping from the retina to the primary visual cortex. Log-polar images have a small central area with maximum resolution (fovea) and a continuously decreasing number of pixels moving toward the periphery. In humans and primates there is the need to move the sensor to take high resolution snapshots of important points in the environment. Likewise, in order to acquire information from the environment, the robot has to move the cameras and place the fovea at interesting locations in the visual scene possibly according to the task at hand. In other words there is the need to have a module that selects information for further visual processing.

In addition, another important requirement for a visual system apt to guide manipulation is that of segmenting objects from a possibly cluttered background. That is, we need both the localization of the object and its segmentation. The problem of segmentation is directly related to the problem of defining what an object is, that is to define which properties distinguish an object from the background.

Our definition of "objecthood" is created in two steps:

- the selection of a set of visual features that combined appropriately can characterize any object of a certain set and allow to segment it from the background;
- the selection of a criterion for grouping features that segment and identify the objects uniquely;

which means, in practice, that we selected certain features and a method for deciding when a specific feature belongs to an object. The features we chose for our implementation are *colored blobs* while the criterion is a consequence of the action of the robot onto objects. In fact, by grasping an

object the robot has the possibility of observing it at will from many points of view and likely with different backgrounds. It is consequently easy to imagine a procedure that selects the constant features across consecutive views of the same object.

By selecting blobs as features we do not propose to define what is an object directly, but rather to consider a sort of “proto-objects”. They are a step above the mere features (e.g. edges), possessing some but not all the characteristics of an object; “proto-objects” in this view are clusters of points on the image “naturally” grouped together. The idea of proto-objects has its roots in psychological [14] and neurobiological literature. In fact it has been proposed that the synchronization of visual cortical neurons can be the carrier of the perceptual grouping phenomenon [15], [16].

In our implementation, we decided that the grouping acted on color and intensity information and thus the grouping of elementary feature leads to the extraction of colored blobs as we mentioned earlier. To simulate the results of the process of grouping, we employed the watershed transform (rainfalling variant) [17] on the edge map resulting from a preliminary feature extraction stage. As a consequence the image is segmented in regions of constant color or a constant gradient of color. A segmentation of this type has been demonstrated to happen in humans before the attention is deployed to the scene [18].

Further, following our definitions, the identity of an object cannot be known without active manipulation, unless some other prior knowledge is inserted into the system. One possible route for learning something about an object autonomously is by means of action. The robot can go beyond the concept of proto-objects, learning a model of any object that is manipulated. In particular, objects are seen here as a collection of proto-objects and their spatial relations. Our solution goes by allowing the robot to manipulate objects and acquire different views of them, and, using the probabilities of occurrence, calculate the probability the collection of blobs being currently fixated is one of the objects the robot is searching for. Then, using these same probabilities, the figure-ground segmentation can be attempted. The complete description of the visual attention model and segmentation procedure can be found in [19].

The segmentation mask is used together with a binocular disparity estimation algorithm which can be used to extract three-dimensional information about the object. The mask defines a region of interest around the object, where a depth map is estimated, and eventually the orientation in space of the object can be extracted and used to guide the behavior of the robot.

In order to achieve a good detection of the object orientation, we developed a fast and robust binocular disparity estimation algorithm, which can work in real-world conditions. The algorithm is based on the work by Van Meerbergen et al. [20], where, given a scanline, all the possible matches between pixels are analyzed by exploring a graph (using dynamic programming) built by assigning a cost to each position pair and each occlusion. The algorithm works at nearly frame rate

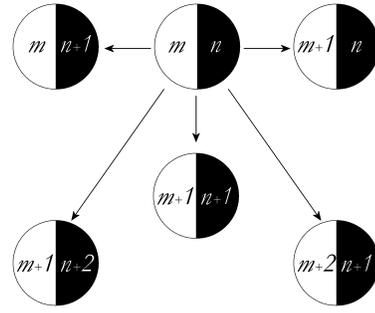


Fig. 3. A section of the graph: the columns have constant disparity ($m - n$).

especially when running on a small portion of the image (a pair of images).

With respect to the original formulation of the algorithm we relaxed the hypothesis of similar extension (over the scanline) of the pixel patterns: something that makes sense when the cameras are quasi-parallel, but on the other hand fails dramatically when the cameras are converging and the objects are very close compared to the interocular distance. The consequences in this last conditions are that a surface can be very different in shape between the two images of the stereo pair. As a result of our changes a short sequence of pixels in one image can match a long one in the other image, which is, as we said, reasonable in the case of our robot.

Each node of the graph represents a pixel pair, one on the left scanline (L_x) and one on the right one (R_y), starting from the nodes containing either (L_0) or (R_0) or both (the first pixels of the scanline) and ending with the nodes containing L_N , R_N or both, where N is the length in pixels of the scanline. Each node is connected with all the following nodes according to these rules:

- 1) *Disparity Range*: $\delta_{min} \leq x - y \leq \delta_{max}$
- 2) *Ordering*: Given another pair ($L_{x'}, R_{y'}$), if $x' \geq x$ then $y' \geq y$, assuming that there cannot be duplicate nodes
- 3) *Continuity*: the nodes following (L_x, R_y) are all the nodes containing (L_{x+1}) or (R_{y+1}) that respect the previous two constraints

Each arc has an associated cost:

$$c = |lum(L_x) - lum(R_y)| + k\beta + \alpha$$

where lum represents the luminance of a pixel, β is a cost (linear) associated to disparity jumps (if a node has disparity δ_i and the following one has disparity δ_j , then $k = |\delta_i - \delta_j|$) and α takes into account the differences along the vertical dimension (to penalize large discontinuities in the final disparity map). Once the graph is built, the minimum cost to traverse it from beginning to end is found using dynamic programming. The disparity map is then constructed by considering the pairing of pixels along the minimum cost path. An example of the construction of the graph when considering two successive levels is shown in figure 3.

Since the complexity of dynamic programming, for each graph (i.e. each scanline), is $O(m)$ – where m is the total

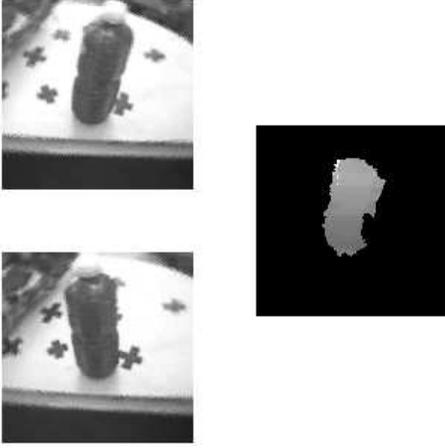


Fig. 4. The Disparity Algorithm Left: The Stereo Pair, Right: The Final Disparity Map, masked. The mask image comes from the attention algorithm.

number of arcs in the graph and it is proportional to the length of the scanlines – we considered only the portion of the image around the object of interest, segmenting the object itself by using the information coming from the saliency algorithm. This reduces both m and the total number of lines to be processed. To increase the robustness, once computed the disparity map D_{l-r} (displacement of the pixels in the left image compared to the ones in the right one), we used it to detect the object position on the right image, according to the formula:

$$D_{l-r}(x) = D_{r-l}(x - D_{l-r}(x)) \quad (1)$$

This result is then used to segment the object on the right image. The images are swapped (left with right) and the disparity is computed again. This new estimate is finally used to validate and correct the previous one. An example of a masked depth map is shown in figure 4.

IV. THE BODY

Humans become skillful at controlling their own body after a long period of postnatal development and probably thousands of trials. As we discussed in the introduction, however, motor development is extremely important and enables the correct perceptual development of the child. For this reason the robot spends the first phase of its artificial development learning how to correctly control the head and the arm to perform various tasks such as visual tracking and reaching for a visually identified target.

Control of the body requires implicit knowledge of its structure (relative position of the limbs, their size, etc.) as well as its dynamical characteristics (e.g. the weight of the body segments). The ensemble of this knowledge is called *body-schema*; experiments in neuroscience have given support to the existence of a body-schema in the primate brain [21], [22]. Graziano and co-workers have found neurons in the motor cortex of the monkey, which code the position of the hand in the visual field.

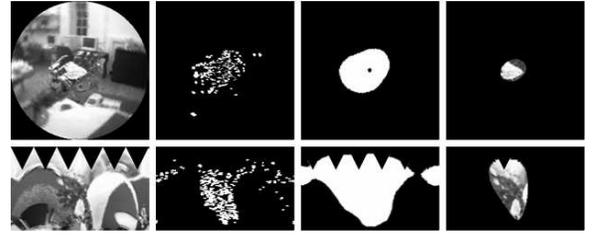


Fig. 5. Hand segmentation: an example.

On the other hand, developmental psychologist have been trying to understand the mechanisms which allow the brain to acquire such a representation. As roboticists we are interested in the same mechanisms as they allow the system (biological or artificial) to autonomously acquire and maintain all parameters required to the control of action and avoid manual estimation and calibration. For this reason the problem has been studied by many authors [1], [23], [24].

We follow here an approach similar to the one of [1], [24] where repeated self-generated actions are exploited for learning. We programmed the robot to perform a periodic movement of the wrist. This motion is observed by the robot visually and *motorically*. In the former case the visual motion is computed by image differencing with an adaptive model of the background. In the latter case the robot computes the first derivative of the encoder feedback. The period of motion of each pixel in the motion image is compared to that of the encoder. Pixels whose motion is periodic and whose period matches that of the joints are selected and grouped together to form the segmentation of the hand. Figure 5 shows an example of the result of this procedure.

We used this procedure to train a neural network to estimate the position of the hand in the visual field given the current robot posture (arm and head joint configuration). Another neural network learns the approximate shape and orientation of the hand given the same segmentation information. Figure 6 show the result of this procedure.

The role of vision during reaching is still debated [25], although experimental results suggest that the sight of the hand is not required for children to start reaching for an object [26], [27] and it is used only relatively late in development to actually adjust the trajectory of the hand during action [28].

Sight of the hand, however, might be used to acquire eye-hand coordination. By tracking the hand the robot builds a mapping between the position of the arm and the corresponding head configuration when fixation is achieved. The hypotheses is that reaching starts by first fixating the object; in this condition the fixation point coincides with the target and uniquely identifies its position with respect to the body. The arm motor command can be obtained by a transformation between the head and arm joint angles, that is by mapping motor variables into motor variables:

$$q_{arm} = f(q_{head}) \quad (2)$$

where q_{arm} and q_{head} are head and arm posture respectively.

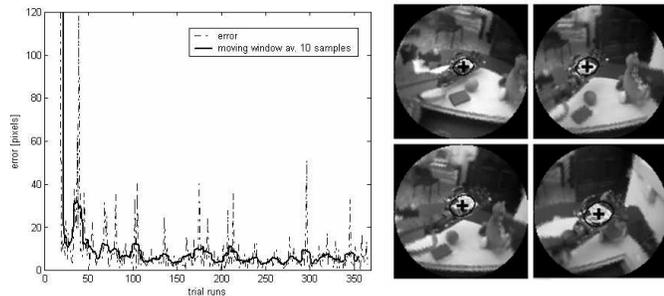


Fig. 6. Learning to localize the hand in the visual field. The convergence of the approximation error as result of training of the neural network (left) and a few exemplar frames of the subsequent prediction (right).

This mapping implicitly implements the inverse kinematics of the arm and it can be learnt when the robot looks at its hand: that is, the robot can maintain the fixation of the hand while moving the arm randomly in the workspace. Every time the arm stops and the eyes have achieved a stable fixation of the hand, a new pair of arm-head positions is acquired and used as a training sample to the neural network approximating the mapping of equation 2. The robot uses the mapping to reach for visually identified objects as soon as a few samples are acquired.

The actual trajectory is computed by linearly interpolating the motor command and the current joint position. The trajectory results in a set of small changes that are effected by a PD controller with gravity compensation. A procedure to learn the gravity component is explained in [12].

V. INTERACTION

We present here a grasping behavior based on the modules described in the previous sections. The interaction with the environment starts when an object is placed in the robot’s hand; the robot detects the object by using the tactile sensors on the palm (see figure 7 frame 1). When pressure on the palm is sensed the fingers close in a stereotyped grasping action. The intrinsic elasticity of the hand (see section II) facilitates grasping, because the fingers automatically adapt to the shape of the object. The robot starts the exploration of the object by bringing it close to the cameras in four different positions and orientations (frames 2-3). During the exploration the robot keeps fixation on the object by tracking the hand. At each position a few frames are acquired and processed as explained in section III to train the model of the object. As the exploration is completed the object is dropped on the table. The robot exploits now the visual model of the object to search for it again (meanwhile the object might have been moved elsewhere by the experimenter) in the visual scene.

The search procedure is driven by a top-down attention module whose contribution exploits the knowledge the robot just acquired about the object. In practice, this happens by selecting the blob whose features better match those of the object’s main blob and performing a saccade motion towards it. After the saccade the object is in the fovea (frame 4 and

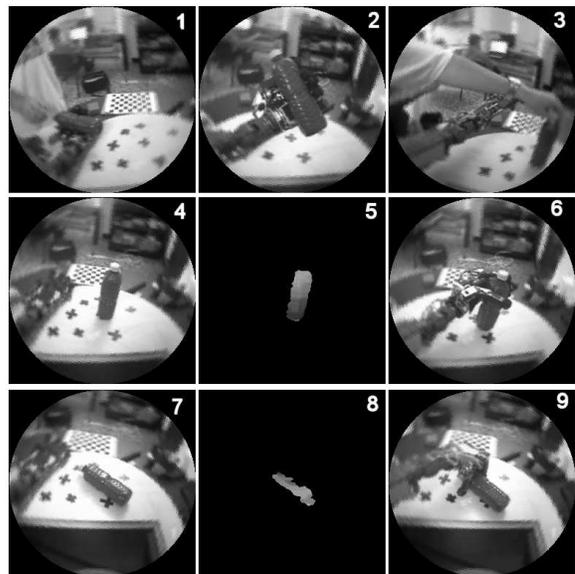


Fig. 7. The execution of a grasping experiment.

7) and its model is matched against the blobs that are now fixated. If the match is positive grasping starts otherwise the search continues. The disparity map of the segmented object is computed to determine the orientation of the object (frames 5 and 8); two different actions are then attempted to maximize the possibility to successfully grasp the object. If the principal axis is oriented horizontally the robot moves the hand above the object, otherwise the hand approaches the object from the side (frames 6 and 9). To determine if the grasping is successful, the robot checks the weight of the object and its “consistence” in the hand (the shape of the fingers around the object). In case of failure another grasping trial is attempted, otherwise the robot waits for a new object to be placed in the palm.

VI. DISCUSSION

We have shown results on two phases of the acquisition of sensorimotor coordination in a upper body humanoid robot. The robot includes a visual attention system employing top-down and bottom-up information. The former is introduced in the system beforehand, whereas the latter is modulated by the robot’s interaction with the environment.

We have shown the importance of the interaction between the environment and the robot for learning. This was demonstrated indirectly, when the robot exploited self-produced actions to explore its own body, and directly when the robot actively explored the visual properties of the objects it grasped.

In the experiment discussed here, we start to link different actions to different objects to investigate the possibility for the robot to autonomously learn which actions are more suitable for different contexts (different objects or environment). Although far from completed, this is meant to enlighten us about the possibility of autonomously enriching the robot’s knowledge of the world. This is not only relevant for action, but

also from a perceptual point of view. Indeed, actions establish a link between events and the causes that have generated them. In other words by acting in the world an “active” agent can link the actions it performs with their consequences. This link can be used in two ways i) for planning, to select the particular action required to bring about a desired consequence and ii) for interpretation, to understand the meaning of an attended event. In the first case, the advantage to use such a representation is that, sometimes, it might be convenient to express goals in perceptual terms. For example pushing an object in a particular direction can be represented by means of the resulting visual motion in the image plane [2]. In the second situation, the only available information is the sensorial experience associated with the event. In this case the robot can search its own experience for an event that closely matches what is observed and can select the action(s) that generated those perceptual experience. For example the sound of an object that hits the floor can be associated with the action of dropping it. Either problems, planning and interpretation, are interesting and challenging, and luckily the solution to both appears to be tightly intertwined with sensorimotor development.

It is fair to say that the system developed so far, although complex, still manifests a certain degree of brittleness perhaps associated to the amount of “handcrafted” components we were nonetheless forced to use to reach this level of functioning in a reasonable amount of time. For instance, the choice of color blobs as features clearly limits the visual system in a way that sometimes prevents the robot from perceiving certain object characteristics. Also, in some other circumstances the residual error in reaching goes unnoticed to the robot that eventually fails to grasp the object reliably. On the object recognition side, objects composed of only a few blobs are easily mistaken for other blobs in the background since a single color (or certain blob combinations) is clearly not discriminative enough (the background might present similar combinations).

We are aware of many of these limitations and, in fact, our ongoing work is exactly aimed to improving the overall performance of the robot both motorically and perceptually with a particular emphasis on the manipulation abilities we deem fundamental for autonomous development.

ACKNOWLEDGMENTS

This work was supported by European Union grants RobotCub (IST-2004-004370) and ADAPT (IST-2001-371173).

REFERENCES

- [1] G. Metta and P. Fitzpatrick, “Early integration of vision and manipulation,” *Adaptive Behavior*, vol. 11, no. 2, pp. 109–128, 2003.
- [2] P. Fitzpatrick, G. Metta, L. Natale, S. Rao, and G. Sandini, “Learning about objects through action - initial steps towards artificial cognition,” in *Proc. of the IEEE Internat’l Conf. on Robotics and Automation*, Taipei, Taiwan, May 2003.
- [3] C. von Hofsten, “An action perspective on motor development,” *Trends in cognitive sciences*, vol. 8, no. 6, pp. 266–272, 2004.
- [4] E. J. Gibson, “Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge,” *Annual Review of Psychology*, vol. 39, pp. 1–41, 1988.

- [5] E. Bushnell and J. Boudreau, “Motor development and the mind: the potential role of motor abilities as a determinant of aspects of perceptual development,” *Child Dev.*, vol. 64, no. 4, pp. 1005–21, 1993.
- [6] S. J. Lederman and R. L. Klatzky, “Hand movements: A window into haptic object recognition,” *Cognitive Psychology*, vol. 19, no. 3, pp. 342–368, 1987.
- [7] C. von Hofsten, “Eye-hand coordination in the newborn,” *Dev. Psychology*, vol. 18, no. 3, pp. 450–61, 1982.
- [8] C. von Hofsten and L. Ronnqvist, “The structuring of neonatal arm movements,” *Child Dev.*, vol. 64, no. 4, pp. 1046–57, 1993.
- [9] L. Natale, G. Metta, and G. Sandini, “Learning haptic representation of objects,” in *International Conference on Intelligent Manipulation and Grasping*, Genoa, Italy, July 2004.
- [10] L. Natale, F. Orabona, G. Metta, and G. Sandini, “Exploring the world through grasping: a developmental approach,” in *Proceedings of the 6th CIRA Symposium*, Espoo, Finland, June 2005.
- [11] E. Torres-Jara, L. Natale, and P. Fitzpatrick, “Tapping into touch,” in *Fifth International Workshop on Epigenetic Robotics (forthcoming)*. Nara, Japan: Lund University Cognitive Studies, July 22–24 2005.
- [12] L. Natale, “Linking action to perception in a humanoid robot: a developmental approach to grasping,” Ph.D. dissertation, University of Genoa, Genoa, Italy, 2004.
- [13] G. Sandini and V. Tagliasco, “An anthropomorphic retina-like structure for scene analysis,” *Computer Vision, Graphics and Image Processing*, vol. 14, no. 3, pp. 365–372, 1980.
- [14] Z. Pylyshyn, “Visual indexes, preconceptual object, and situated vision,” *Cognition*, vol. 80, pp. 127–158, 2001.
- [15] R. Eckhorn, R. Bauer, W. Jordan, M. Brosch, M. Kruse, W. Munk, and H. J. Reitboeck., “Coherent oscillations: A mechanism of feature linking in the visual cortex?” *Biological Cybernetics*, vol. 60, pp. 121–130, 1988.
- [16] C. M. Gray, P. Knig, A. K. Engel, and W. Singer, “Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties,” *Nature*, vol. 338, pp. 334–336, 1989.
- [17] P. D. Smet and R. Pires, “Implementation and analysis of an optimized rainfalling watershed algorithm,” in *IS&T/SPIE’s 12th Annual Symposium Electronic Imaging 2000*, San Jose, California, USA, 2000, pp. 759–766.
- [18] J. Driver, G. Davis, C. Russell, M. Turatto, and E. Freeman, “Segmentation, attention and phenomenal visual objects,” *Cognition*, vol. 80, pp. 61–95, 2001.
- [19] F. Orabona, G. Metta, and G. Sandini, “Object-based visual attention: a model for a behaving robot,” in *3rd International Workshop on Attention and Performance in Computational Vision at the CVPR 2005*, San Diego, CA, USA, 2005.
- [20] G. V. Meerbergen, M. Vergauwen, and M. Pollefeys, “A hierarchical symmetric stereo algorithm using dynamic programming,” *International Journal on Computer Vision Graphics and Image Processing*, vol. 47, no. 1/2/3, pp. 275–285, 2002.
- [21] M. Graziano, “Where is my arm? the relative role of vision and proprioception in the neuronal representation of limb position,” *Proceedings of the National Academy of Science*, vol. 96, pp. 10 418–10 421, 1999.
- [22] M. Graziano, D. Cooke, and C. Taylor, “Coding the location of the arm by sight,” *Science*, vol. 290, pp. 1782–1786, 2000.
- [23] Y. Yoshikawa, K. Hosoda, and M. Asada, “Does the invariance in multi-modalities represent the body scheme ? - a case study with vision and proprioception -,” in *2nd Intelligent Symposium on Adaptive Motion of Animals and Machines*, Kyoto, Japan, 2003.
- [24] P. Fitzpatrick and A. Arsenio, “Feel the beat: using cross-modal rhythm to integrate perception of objects, others and self,” in *Fourth International Workshop on Epigenetic Robotics*, vol. 117. Genoa: Lund University Cognitive Studies, 2004.
- [25] J. Saunders and D. Knill, “Humans use continuous visual feedback from the hand to control fast reaching movements,” *Experimental Brain Research*, vol. 152, pp. 341–352, 2003.
- [26] R. Clifton and M. C. D.W. Muir, D.H. Ashmead, “Is visually guided reaching in early infancy a myth?” *Child Dev.*, vol. 64, no. 4, pp. 1099–110, 1993.
- [27] R. Clifton, P. Rochat, D. Robin, and N. E. Berthier, “Multimodal perception in the control of infant reaching,” *J Exp Psychol Hum Perform.*, vol. 20, no. 4, pp. 876–86, 1994.
- [28] D. Ashmead, M. McCarty, L. Lucas, and M. Belvedere, “Visual guidance in infants’ reaching toward suddenly displaced targets,” *Child Dev.*, vol. 64, no. 4, pp. 1111–27, 1993.