

## VISUOMOTOR MAPS FOR ROBOT TASK LEARNING THROUGH IMITATION

Raquel Frizera Vassallo <sup>\*,1</sup> José Santos-Victor <sup>\*\*</sup>  
Hans Jörg Schneebeli <sup>\*</sup>

<sup>\*</sup> Universidade Federal do Espírito Santo, Vitória, ES,  
Brazil - {raquel,hans}@ele.ufes.br

<sup>\*\*</sup> Instituto de Sistemas e Robótica - Instituto Superior  
Técnico, Lisboa, Portugal - {raquel,jasv}@isr.ist.utl.pt

**Abstract:** We propose an approach that allows a robot to learn a task through imitation, using motor representations, as suggested by recent findings in neuroscience. The robot relies on a visuomotor map to convert visual information into motor data. Then, by observing and imitating other agents, the robot can learn a set of elementary motions (motor vocabulary), that will eventually be used to compose more complex actions, for each specific task domain. We illustrate the approach in a mobile robotics task. Egomotion estimation is used as a visuomotor map, that allows the robot to learn a motor vocabulary for topological mapping and navigation. The approach can be extended to different robots and applications. Encouraging results are presented and discussed.

**Keywords:** Learning, Robot Navigation, Egomotion, Omnidirectional Vision

### 1. INTRODUCTION

In the 1960's, robots were almost exclusively used in industries: robot arms used to automate the workspace, achieving high precision and repeatability on the assembly line for mass production. No "intelligence" was involved and everything was programmed a priori, enabling a robot to repeat the same set of movements for weeks or months. Even in the absence of sensory information, robots would still accomplish their tasks.

Since then, several new markets for robotics have emerged, calling for new technologies and paradigms. Nowadays, there is a move towards applications where the environment is much less structured, tasks that require a higher degree of sensing and interacting with humans and other agents (toys, house keeping, elderly and patient assistance, etc). Such tasks pose a great demand for more complex, intelligent and autonomous robots. They should be flexible systems, able to learn in an open-ended way, and adapt while interacting with other agents, people and environment.

The challenges that rise from these new demands are how to program such complex systems and

achieve flexibility. Nature seems to have already solved this problem. Animals are complex and flexible systems, able to learn and adapt through a simultaneous motor, sensorial and cognitive development. Thus, instead of programming all situations, actions and reactions, robots could learn and adapt through a developmental approach also. After acquiring a self-knowledge, that allows knowing their own body and how to control movements, robots could improve and become more flexible by learning through interaction.

One way to learn and afford a high degree of flexibility is through imitation, a powerful social learning and adaptation method (Dautenhahn, 1995). Imitation avoids undergoing through extensive trial and error, since the imitator learns directly from the teacher's experience. Learning by imitation has been addressed by many researchers and there are interesting results for task learning, skill acquisition and communication (Billard, 2002). Other works in robotics inspired by imitation in animals can be seen in (Dautenhahn and Nehaniv, 2002).

Recent findings in neuroscience have shed new insight on the way imitation might be achieved in nature. A special group of visuomotor neurons,

---

<sup>1</sup> Partially supported by the portuguese Fundação para a Ciência e Tecnologia (FCT), Programa Operacional Sociedade de Informação (POSI) in the frame of QCA III, EU-Project IST-2000-28159 MIRROR and CAPES, Brazil.

designated *mirror neurons*, was discovered in area F5 of monkey’s brain (Fadiga *et al.*, 2000). These neurons activate either when the monkey performs an action or when it sees the same action performed by a demonstrator or another monkey. Usually related to hand and mouth movements, they seem to have the ability to recognize gestures and constitute the basis of the mechanism of action imitation and understanding (Rizzolatti and Arbib, 1998).

The visual activation of mirror neurons illustrates the relationship between perception and action (Lopes and Santos-Victor, 2003). The fact that mirror neurons are located in the motor area of the brain, suggests that observed actions are first converted to motor information, before being recognized. Usually, recognizing actions in visual space is more difficult and costly than in motor space, where it becomes independent of frame position, orientation or view-point.

The existence of a visuomotor mapping that converts visual information into motor measurements is also supported by biology. When newborns look to their hands and own movements, they are probably learning the relationship between motor action and visual stimuli (Metta *et al.*, 2001). Through this visuomotor mapping, children become able to recognize and repeat movements.

We propose a methodology for open-ended robot adaptation, using learnt motor representations to acquire capabilities that go beyond initial programming. A central role is played by a visuomotor mapping that allows the robot to relate visual stimuli to its own motor representations. The advantages of this approach are: (i) through interaction, a robot is able to learn task-specific movements that are not defined a priori; (ii) it represents a natural manner of adapting the application to the robot’s motion capabilities; (iii) it can be applied to different robots and environments. Whenever the application changes, a new set of movements can be taught. The overall idea is illustrated in Figure 1.

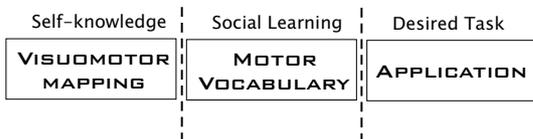


Fig. 1. Robot learning and adaptation approach.

The first step represents a motor self-knowledge: the visuomotor mapping. It allows the robot to learn its motor capability, by observing the visual effects due its own motion. It defines a relation between a motor space (i.e. linear and angular velocities, joint position and orientation) and a visual space (i.e. image point coordinates, optical flow) according to the robot’s body (i.e. wheels

or legs) and visual system (i.e. perspective or omnidirectional cameras, color or B&W). This visuomotor mapping can be computed explicitly through equations or learnt through the observation of robots own movements.

Once it has a visuomotor mapping ability, the robot becomes able to recognize actions while observing and interacting with another agent. Through a social learning method (i.e. imitation), the robot learns a set of purposive actions needed for solving a motor task. This set of elementary actions can be seen as a motor vocabulary internally represented by the robot’s motor repertoire.

Finally, the learnt vocabulary is used to perform more complex actions related to the desired application. This approach seems to be a flexible way of adapting the desired task and the way it should be performed to the robot’s motion repertoire. Different robots can learn the same motor vocabulary and also for different tasks, different vocabularies can be taught.

As an example, we implemented a motor task for a mobile robot. The visuomotor mapping is done through egomotion estimation while the social learning is based on imitation/following. The learnt motor vocabulary is used for topological mapping and navigation. The reason of choosing such a task was, besides illustrating the proposed methodology, to implement an application different from previous works done on imitation.

In Section 2 we present the implemented visuomotor mapping. In Sections 3 and 4 the motor vocabulary learning and the mobile robot application are described. Some experiments and results are listed in Section 5 while our conclusions and future work are discussed in Section 6.

## 2. EGOMOTION AS VISUOMOTOR MAPPING

A visuomotor map must convert visual measurements to motor space so the robot can relate what it sees with how it moves. Egomotion estimation can be seen as a visuomotor mapping method. It recovers motor information from optical flow measurements obtained from a sequence of images.

In this work, egomotion estimation was done using omnidirectional images. The problem becomes easier if a spherical motion field is used instead of a planar field obtained with perspective cameras (Nelson and Aloimonos, 1988). From a sequence of omnidirectional images, we first calculated optical flow and then remapped image flow vectors to the unit sphere surface through an image Jacobian matrix. On such hemispherical motion field either the focus of expansion (FOE) or the focus of contraction (FOC) are visible. Finally, motor

information was estimated from the motion field adapting an egomotion algorithm designed for planar projection to spherical projection (Vassallo *et al.*, 2002a).

Previously, the Jacobian matrix needed to remap image flow vectors was defined according to the system projection model (Gluckman and Nayar, 1998). Instead, we used a general projection model defined by (Geyer and Daniilidis, 2000) to define a general Jacobian. This projection model can represent different omnidirectional systems by combining a mapping of a 3D point  $P$  to a sphere followed by a projection to the image plane. The center of the sphere  $C(0,0,0)$  lies on the optical axis of the projection to the plane and represents the origin of the reference frame (see Fig. 2).

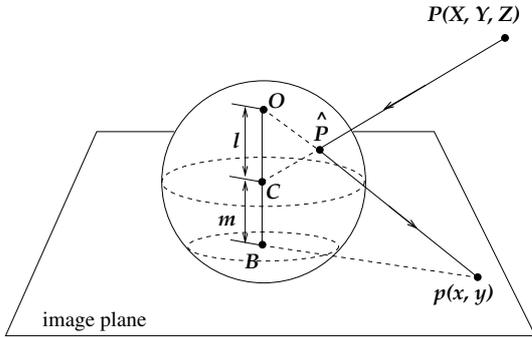


Fig. 2. The general projection model.

The parameters  $l$  and  $m$  adjust the model to different systems and correspond to the distances from the sphere center  $C$  to the projection center  $O$  and to the projection plane.

By back-projecting an image point  $(x, y)$ , we can obtain a point on the unit sphere  $\hat{P}(\hat{X}, \hat{Y}, \hat{Z})$  corresponding to the direction of the incoming ray from the original 3D point  $P(X, Y, Z)$  (see Eq. 1).

$$\begin{bmatrix} \hat{X} \\ \hat{Y} \end{bmatrix} = \frac{lA + \text{sign}(A)\sqrt{(x^2 + y^2)(1 - l^2) + (A)^2}}{x^2 + y^2 + (A)^2} \begin{bmatrix} x \\ y \end{bmatrix} \quad (1)$$

$$A = l + m \quad \hat{Z} = \pm \sqrt{1 - \hat{X}^2 - \hat{Y}^2}$$

where  $\hat{Z}$  becomes negative if  $|l+m|/l > \sqrt{x^2 + y^2}$  and positive otherwise. Neither the camera intrinsic parameters, image center and focal length are considered in the above expression.

To reproject flow vectors from the image plane to the sphere surface, the general *Jacobian* matrix  $J$  is defined by differentiating the spherical coordinates  $(\hat{X}, \hat{Y}, \hat{Z})$  on the back-projection equation with respect to the image coordinates  $(x, y)$  (Vassallo *et al.*, 2002a). It maps image velocity vectors to the unit sphere surface, transforming a planar flow field to a hemispherical motion field that will help estimate egomotion (see Eq. 2 and Fig. 3).

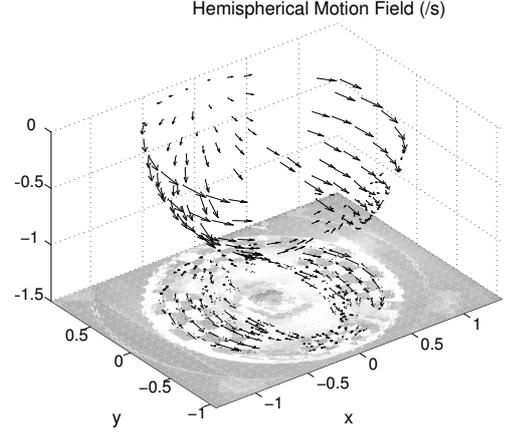


Fig. 3. Image velocities remapped to the unit sphere surface.

$$\begin{bmatrix} \dot{\hat{X}} & \dot{\hat{Y}} & \dot{\hat{Z}} \end{bmatrix}^T = J \begin{bmatrix} \dot{x} & \dot{y} \end{bmatrix}^T \quad \text{with} \quad J = \begin{bmatrix} \frac{\partial \hat{X}}{\partial x} & \frac{\partial \hat{Y}}{\partial x} & \frac{\partial \hat{Z}}{\partial x} \\ \frac{\partial \hat{X}}{\partial y} & \frac{\partial \hat{Y}}{\partial y} & \frac{\partial \hat{Z}}{\partial y} \end{bmatrix}^T \quad (2)$$

Egomotion was estimated by adapting the Bruss and Horn (Bruss and Horn, 1983) algorithm designed for planar perspective projection to spherical projection. The motion field  $U$  at a point  $\hat{P}$  on the unit sphere is a function of the camera rotation  $\Omega$ , translation  $T$  and the corresponding 3D point depth  $R = \sqrt{X^2 + Y^2 + Z^2}$  (Eq. 3).

$$U(\hat{P}) = \frac{1}{R}((T \cdot \hat{P})\hat{P} - T) - \Omega \times \hat{P} \quad (3)$$

Depth dependency is removed by taking the cross product with  $\hat{P}$  and the dot product with  $T$  (see Eq. 4). Estimation was done through an iterative process using non-linear minimization considering  $|T| = 1$ , once it is not possible to recover the linear velocities values.

$$T \cdot (\hat{P} \times (U + (\Omega \times \hat{P}))) = 0 \quad (4)$$

Egomotion acts as a visuomotor mapping that converts visual information to motor measurements,  $T(T_x, T_y, T_z)$  and  $\Omega(\omega_x, \omega_y, \omega_z)$ , according to the robot's motor and visual capabilities.

### 3. LEARNING A PURPOSEFUL MOTOR VOCABULARY THROUGH IMITATION

Through imitation, a robot can get the same type of visual information used by the visuomotor mapping to perceive and define an internal motor coding for the executed actions. Elementary movements can be learnt and constitute a motor vocabulary important for performing a particular motor task. This approach allows a flexible way of adapting the task to the available robot's motor capability.

In this work, we considered a mobile robot which motion is restricted to the ground plane. Imitation was implemented as a *person-following* behavior. For simplicity, we assumed that the person to be followed carries a distinctive green-colored rectangle. The target is first detected using the hue channel of frontal images captured by a color camera. Noise in the resulting binary image is filtered through morphological operators and the largest remaining blob is selected. The contour is detected and the rectangle lines are estimated by a robust fitting procedure. Finally, the corners coordinates are determined from the lines intersection, as shown in Figure 4. A visual servoing strategy was implemented so that the robot could follow the green rectangle at a predefined distance (1m) and oriented toward its center (Vassallo *et al.*, 2002b).

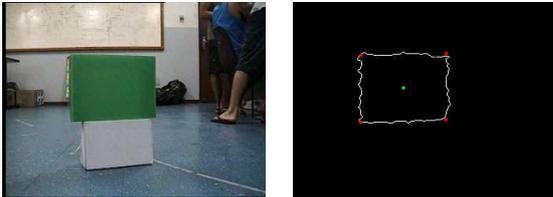


Fig. 4. The green rectangle's corners detection.

While the robot follows a person, it uses its visuomotor mapping (egomotion) to perceive the movements. Then all the egomotion estimations are classified into clusters by an unsupervised learning method based on K-means. The number of wanted clusters is defined by the user. Centroids represent the learnt movements that constitute a motor vocabulary. Labels can also be associated to each movement as they were motor words.

Once the vocabulary is created, the robot can recognize movements of interest and use them to perform a desired task. Movement recognition is done in motor space using the Euclidean distance as the discriminant function.

#### 4. THE TASK: TOPOLOGICAL MAPPING AND NAVIGATION

Egomotion was considered as the visuomotor mapping for a mobile robot and a following behavior was used to teach the robot a purposive motor vocabulary. The learnt vocabulary was then used for a specific application: building topological maps and navigation.

For map building, the robot was guided through the environment. During motion, the robot captures omnidirectional images to define nodes and recognizes movements using the learnt motor vocabulary to associate motor words to links in the map. The decision of whether or not inserting a new node in the map is taken based on a

comparison with the previously stored reference image or changes in motion. The *sum of squared differences* (SSD) is used as a metric to assess the difference between images. Whenever a new node is stored, the most frequently recognized motor word is attributed to the link between that node and the previous one.

To navigate using the created map, the robot first auto-localize and find a path between the initial and the goal positions. During map navigation, it monitors its progress along the route by comparing the captured images against the current image node and the subsequent one in the path. A hill-climbing strategy is used to determine when the robot position should be updated.

Whenever a new node is reached, the motor word stored in the following link determines the next motor command. This behavior continues until the final location is reached. Some experiments and results are shown in the next section.

## 5. EXPERIMENTS AND RESULTS

The robot we used is a Pioneer DX2 equipped with an on board computer (Pentium II MMX - 266 MHz - 128 MRAM). A color camera is used for the following behavior and a B&W catadioptric omnidirectional camera is used for egomotion and mapping. The omnidirectional system is mounted on the top of the robot with its axis coincident to the platform's rotation axis (see Figure 5).

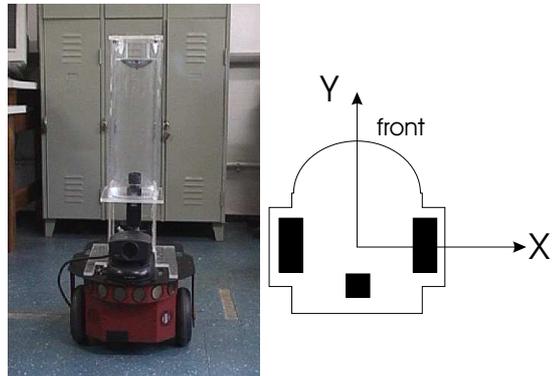


Fig. 5. The robot, the vision systems and the adopted reference frame.

Some egomotion results (vectors  $(T_x, T_y)$  and  $\omega_z$ ) obtained by a mobile robot while moving on the ground plane and observing its own motion are shown in Figure 6. Estimation values and errors calculated by comparing the results with the nominal values are also listed in Table 1.

During the vocabulary learning phase, a set of following experiments were done. Egomotion values were normalized and classified into clusters defining the motor words. Because the robot is a differential platform, just the angular velocity  $\omega_z$

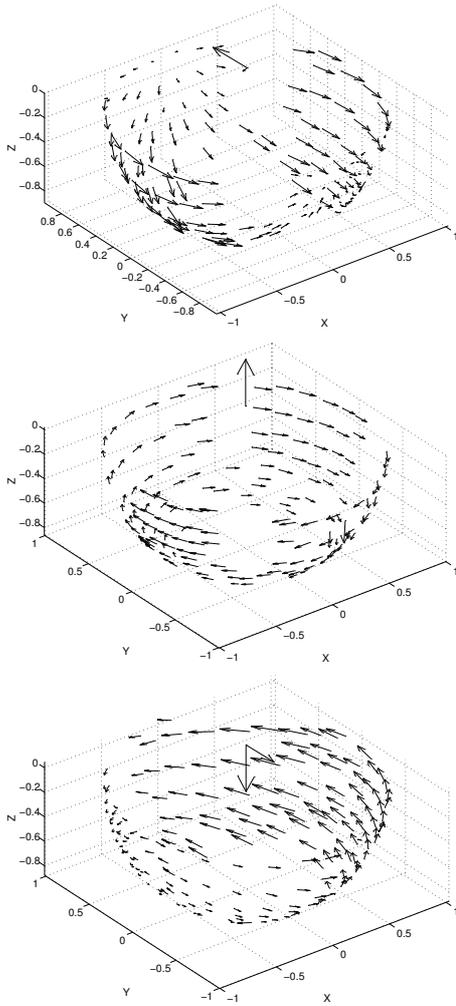


Fig. 6. Examples of (a) Translation, (b) Rotation and (c) Combined move. Egomotion vectors  $(T_x, T_y)$  and  $\omega_z$  indicated in the middle of the hemispherical flow.

(a)	Translation $T (T_x \ T_y)$	Rotation $\Omega (\omega_z) (^\circ/s)$
nominal	$T = [0 \ 1]$	$\Omega = [0]$
estimation	$\hat{T} = [0.0128 \ 0.9999]$	$\hat{\Omega} = [-0.0132]$
error	$e_t = 0.734^\circ$	$e_{\Omega} = [-0.0132]$
(b)	Translation $T (T_x \ T_y)$	Rotation $\Omega (\omega_z) (^\circ/s)$
nominal	$T = [0 \ 0]$	$\Omega = [-3.1255]$
estimation	$\hat{T} = [0 \ 0]$	$\hat{\Omega} = [-3.4965]$
error	$e_t = 0^\circ$	$e_{\Omega} = [-0.371]$
(c)	Translation $T (T_x \ T_y)$	Rotation $\Omega (\omega_z) (^\circ/s)$
nominal	$T = [0 \ -1]$	$\Omega = [3.8776]$
estimation	$\hat{T} = [0.0252 \ -0.9997]$	$\hat{\Omega} = [4.3284]$
error	$e_t = 1.45^\circ$	$e_{\Omega} = [0.4508]$

Table 1. Egomotion estimations/errors.

and the  $T_y$  component of translation corresponding to the robot forward direction were considered. Although both  $T_x$  and  $T_y$  were estimated, the values for  $T_x$  are usually due to sliding of the robot wheels or noise in the estimation process.

The clusters found in the  $(T_y, \omega_z)$  space are shown in Figure 7. The mean values are represented by asterisks and Voronoi lines separating the various

clusters are drawn. The centroids represent the purposive motor vocabulary. Names were given to each cluster defining motor words which are detailed in Table 2.

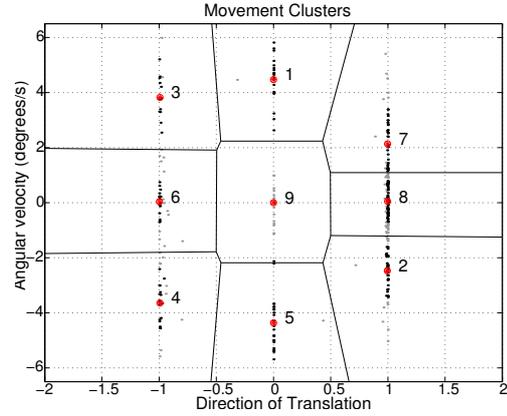


Fig. 7. Clusters of the created vocabulary. Black/gray points indicate inliers/outliers.

N	Motor Word	$T_y$	$\omega_z$ (degrees/s)
W 1	left turn	0	4.4722
W 2	front right	0.9948	-2.4749
W 3	back left	-0.9927	3.8223
W 4	back right	-0.9927	-3.6440
W 5	right turn	0	-4.3646
W 6	back	-0.9972	0.0386
W 7	front left	0.9973	2.1350
W 8	front	0.9974	0.0488
W 9	stopped	0	0

Table 2. Motor words for each cluster.

The vocabulary in Table 2 was created with a specific intention: topological mapping and navigating. The executed path for map building is shown in Figure 8. The places where images were captured are indicated by asterisks.

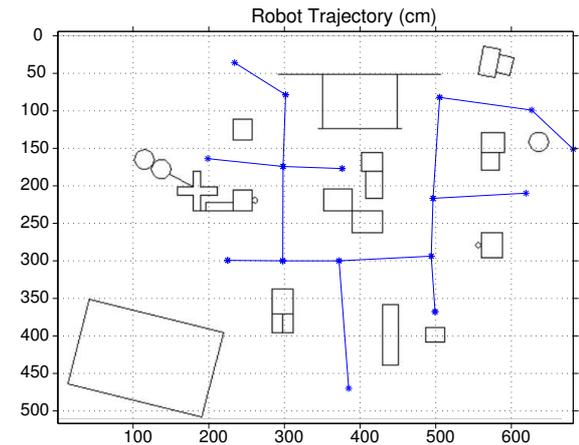


Fig. 8. Robot trajectory (odometry) during map building. Asterisks indicate map nodes.

Some navigation experiments are shown in Figure 9. The robot was asked to navigate between different points crossing the map and moving either in the same or opposite direction of that used for map creation.

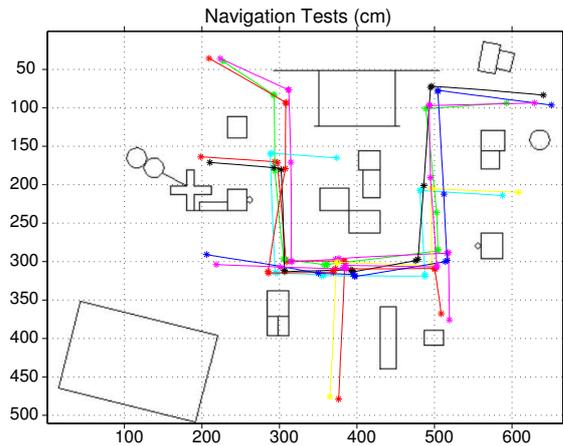


Fig. 9. Robot trajectories (odometry) when map navigating. Asterisks are position updates.

The places where the robot updated its position are indicated by asterisks. Although some points correspond to the same node in the map, the asterisks did not happen exactly at the same coordinates but at the same region, characterizing a qualitative navigation.

## 6. CONCLUSIONS AND FUTURE WORK

We proposed an approach that allows a robot to learn a task through imitation, using motor representations. Inspired by biology and brain theory, a visuomotor mapping is considered to convert visual information into motor commands allowing the robot to relate what it sees with how it moves and to recognize movements while imitating another agent.

The main contributions are (i) to make a robot able to learn through interaction sets of movements that were not programmed or defined a priori and (ii) provide a natural manner of adapting the desired application and the way it is performed to the robot's motion repertoire. This approach represents an open-ended methodology for robot adaptation that can be applied to different robots and applications. The visuomotor mapping and the created motor vocabulary are defined according to robot's body and visual system. Also, whenever the application or environment changes a new set of movements can be taught.

We illustrated the approach by a mobile robot that used egomotion as the visuomotor mapping and that learnt, through a following behavior, a motor vocabulary used for topological mapping and navigating. Omnidirectional images improved both egomotion and mapping/navigation processes, once the large field of view helps finding and tracking objects and allows more of the environment to be caught in just one image.

The obtained results are encouraging. Our future work will focus on extending this ideas to other

robots/applications and implementing a way of learning the visuomotor mapping instead of computing it explicitly through egomotion equations.

## REFERENCES

- Billard, Aude (2002). *Imitation in Animals and Artifacts*. Chap. 11 - Imitation: A Means to Enhance Learning of a Synthetic Protolanguage in Autonomous Robots, pp. 281–310. The MIT Press.
- Bruss, A. and B. K. P. Horn (1983). Passive navigation. *Computer Vision, Graphics and Image Processing* **21**, 276–283.
- Dautenhahn, K. (1995). Getting to know each other - artificial social intelligence for autonomous robots.. *Robotics and Autonomous Systems* **16**, 333–356.
- Dautenhahn, Kerstin and Nehaniv, Chrystofer L., Eds.) (2002). *Imitation in Animals and Artifacts*. The MIT Press.
- Fadiga, L., L. Fogassi, V. Gallase and G. Rizzolatti (2000). Visuomotor neurons: ambiguity of the discharge or 'motor' perception?. *Int. J. Psychophysiology* **35**, 165–177.
- Geyer, C. and K. Daniilidis (2000). A unifying theory for central panoramic systems and practical implications. In: *ECCV 2000*. Vol. 2. pp. 445–461.
- Gluckman, J. and S. K. Nayar (1998). Ego-motion and omnidirectional cameras. In Proc. of 6th International Conference on Computer Vision - ICCV98.
- Lopes, M. C. and J. Santos-Victor (2003). Visual transformations in gesture imitation: what you see is what you do. In: *IEEE Intl. Conference on Robotics and Automation - ICRA2003*.
- Metta, Giorgio, Giulio Sandini and Lorenzo Natale (2001). Sensorimotor interaction in a developing robot. In: *1st Intl. Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*. pp. 18–19.
- Nelson, R.C. and J. Aloimonos (1988). Finding motion parameters from spherical motion fields (or the advantages of having eyes in the back of your head). *Biological Cybernetics* **58**, 261–273.
- Rizzolatti, G. and M. A. Arbib (1998). Language within our grasp. *Trends in Neurosciences* **21**(5), 188–194.
- Vassallo, R. F., J. Santos-Victor and H.J. Schneebeli (2002a). A general approach for egomotion estimation with omnidirectional images. In: *IEEE Workshop on Omnidirectional Vision 2002 (with ECCV)*. pp. 97–103.
- Vassallo, R., J. Santos-Victor and H. Schneebeli (2002b). Using motor representations for topological mapping and navigation. In: *Intl. Conference on Intelligent Robots and Systems - IROS2002*. pp. 478–483.