A new omnidirectional vision sensor for the Spatial Semantic Hierarchy

Emanuele Menegatti†, Mark Wright§, Enrico Pagello†‡

†Department of Informatics and Electronics
University of Padua
Padua, Italy
§Edinburgh Virtual Environment Center
University of Edinburgh
Edinburgh, UK
‡Institute LADSEB of CNR
Padua, Italy
emg@dei.unipd.it
www.dei.unipd.it/ricerca/airg/

Abstract— ***In this paper we propose a new sensor for the Spatial Semantic Hierarchy created by Benjamin Kuipers. To prove the effectiveness of this new sensor it has been used as a sole sensor for a robot. The task of the robot is to build a topological map of an unknown environment using the Spatial Semantic Hierarchy. In the paper, we present the strict link that it is possible to create between the Spatial Semantic Hierarchy and the omnidirectional images. We propose a set of topological events, that it is possible to identify in the sequence of images acquired while the robot moves. These topological events can be used to pose a discrete set of places that will be the nodes of the topological map. The results of simulated experiments, and the experiments we carried out with a real robot, shows the feasibility of this approach. At the moment we are testing a new multi-part mirror expressively designed for this application.***

I. Introduction

The navigation problem and the map building problem are two of the oldest problems studied by the mobile robotics community [?]. A wide spectrum of solutions have been proposed using the most disparate sensors. In this paper, we propose a new approach to the map building task: the fusion of a well understood method, the Spatial Semantic Hierarchy (**SSH**) [?] with a relatively new sensor, an omnidirectional vision system [?].

In the last years, omnidirectional vision systems have been exploited successfully in robot navigation and map building. The success of this kind of sensors is explained by the wide field of view achievable. Omnidirectional sensors offer in one shot a global view of the surroundings. Therefore, the robot does not need to look around using moving parts (cameras or mirrors) or turning on the spot [?]. Omnidirectional vision systems have been used more and more in the popular international Robocup competitions (www.robocup.org). Lima used an omnidirectional sensor for the self-localization of the robot in the field of play. In this application, the omnidirectional mirror is designed to give a bird's eye view of the pitch. This permits to exploit the natural landmarks of the soccer field (goals and fields lines) for a reliable self-localization [?]. In another robot team, Asada used a goal-keeper fitted with an omnidirectional vision system with a learning capability. To reduce the learning time, the omnidirectional sensor is fitted with an attention control provided by an active zoom mechanism that permits to select a restrict area of the omnidirectional image [?].

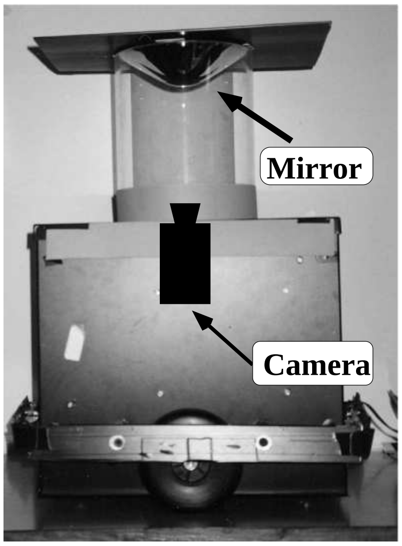


Fig. 1. The robot with its omnidirectional sensor.

The main disadvantage of omnidirectional vision with respect to perspective vision is the poor resolution of the images. In fact, with a catadioptric sensor, the light is gathered from a much wider area than with a perspective camera. In the map building task the low resolution of the omnidirectional images is not a shortcoming. We are more interested in the position of the objects in the environment, than in the details of their surface. However, for particular applications, there might be an interest to observe at higher resolution certain areas around the robot. Within certain limits, it is possible to design mirrors that maximize the image resolution in the most interesting regions of the scene. As we will discuss in Section **??**, at the moment we are working to increase the performance of the robot with a custom designed mirror. We designed a mirror profile that increases the image resolution near the base of the robot [?].

In most systems, the motion of the robot is considered known. In this work we did not use direct information on the movements of the robot. The robot had to "infer" its movements from the vision sensor (on the possibility to estimate the egomotion with omnidirectional vision sensor, see the work of Svoboda [?]).

The aim of this paper is to show that a catadioptric omnidirectional sensor is a good sensor for the SSH. In Section II, we summarise the basics of the SSH, focusing on the concepts exploited by our vision system. In Section III, we present the omnidirectional sensor used. In Section IV, we show the strict link that is possible to create between the SSH and the omnidirectional frame sequences. In Section V, we state the assumptions made in our research. In Section VI, we explain which features and events we decided to extract from the omnidirectional sequences. In Section VII, we present the simulated experiments and the actual experiments we carried out to test our system. In Section **??** we show how it is possible to improve the omnidirectional catadioptric sensor with a new multi-part mirror we designed. Eventually, conclusions are drawn in Section **??**.

II. Spatial Semantic Hierarchy

The Spatial Semantic Hierarchy (**SSH**) is a model of the knowledge of large-scale spaces of humans, intended to serve as a "method for robot exploration and map building" [?]. The SSH is made up of several layers. Each layer can be implemented independently, even if they strongly interact (in accordance with the *Behavior Based Approach* of Brooks). Let us see briefly what each layer is about:

- The *Sensory Level* is the interface with the agent's sensory system. It extracts the useful environmental clues from the continuous flux of information it receives from the robots' sensors.
- The *Control Level* describes the world in terms of continuous actions called control laws. A control law is a function which relates the sensory input with the motor output. A control law is retained until a transition of state is detected. A transition of state can be detected with a function called a **distinctiveness measure**. The distinctiveness function must be identified depending on the sensor, on the features and the events which have to be extracted from the environment.
- The *Causal Level* abstracts a discrete model of the environment from the continuous world. In other words, it is at the causal level that a discrete set of places is extracted from the continuous world. These places will be the nodes of the topological map. The discrete model of the environment is composed of **views** , **actions** and the causal relations between them. A **view** is defined as the sensor's reading at a **distinct** place. A **distinct place** is a place where a transition of state is detected. An **action** is defined as the application of a sequence of control laws. It is convenient to classify actions into two categories: *travels* and *turns*. "A **turn** is an action that leaves the agent at the same place. A **travel** takes the agent from one place to another" [?].
- The *Topological Level* represents an environment as places, paths and regions connected or contained one in the other. To use Kuipers's words:

The topological model of the environment is constructed by the non monotonic process of ***abduction****, positing the minimal set of places and paths needed to explain the regularities observed among views and actions at the causal level.*

- The *Metrical Level* augments the topological representation of the environment by including metric properties such as distance, direction, shape, etc. This may be useful, but is seldom essential.

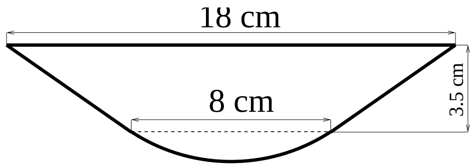


Fig. 2. The mirror profile

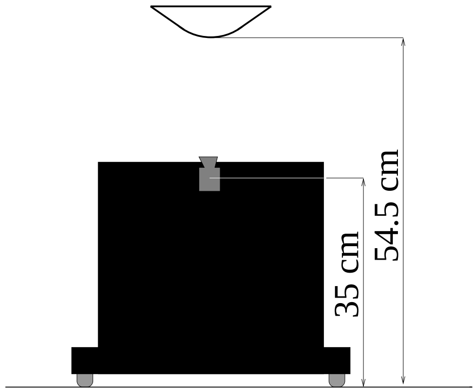


Fig. 3. The vision system

So far, the SSH has only been implemented either on simulated robots or on real robots with very simple sensors (as sonars). As far as we know, no attempt to use an omnidirectional vision sensor has been made. In the following, we will present the omnidirectional sensor used and we will show why an omnidirectional sensor is a good sensor for building a topological map within the Spatial Semantic Hierarchy frame.

III. Omnidirectional Sensor

In this paper we used a mobile robot fitted with an omnidirectional sensor. The robot is depicted in Figure 1. It is the reserve goal keeper of the Azzurra Robot Team

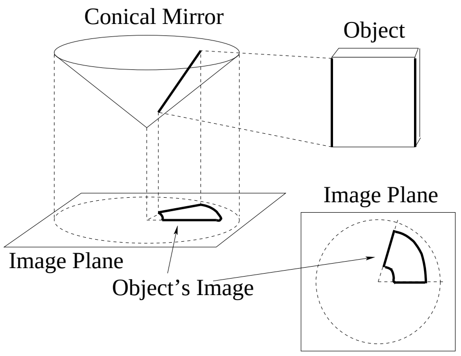


Fig. 4. The conical projection showing how a vertical line is mapped in a radial line in the image plane (adapted from Yagi).

(ART)[1]. Its omnidirectional sensor is composed of a perspective camera pointed upwards at the vertex of a convex mirror.

The optical axis of the camera and the geometrical axis of the mirror are aligned. The mirror is supported by a transparent Perspex cylinder. The shape of the mirror is obtained by the intersection of a cone and a sphere[2] [?]. The dimensions of the cone, the radius of the sphere and the position of the point of intersection are calculated such that the cone and the sphere are tangential in the intersection point. If they were not tangential, a discontinuity would be present in the image. Figure 2 and Figure 3 present the sketches of the mirror profile with its dimensions and the distances of the mirror from the camera and the floor.

Consider Figure 4 to understand how an omnidirectional sensor maps the scene into the image plane[3]. The vertical edges in the scene are mapped in the image plane as radial lines originating from the point corresponding to the tip of the mirror. The azimuth of a radial line in the image corresponds to the azimuth of the vertical edge in the scene, as viewed from the optical axis of the camera. The horizontal lines are mapped in curved lines which shape depends strongly on the geometry of the mirror. Note that the omnidirectional mirrors have a rotational invariance. If the sensor rotates through a certain angle about the vertical, the relative position of the objects in the image does not change. The image is only rotated and all the objects appear to have experienced an azimuthal shift equal to the angle of rotation.

IV. Why omnidirectional vision is good for the Spatial Semantic Hierarchy.

In the introduction we reported the reasons for the success of omnidirectional vision sensor in the map building task. When working within the SSH frame, other benefits of omnidirectional vision come into view.

The omnidirectional images can be strictly correlated with the **views**[4] introduced in the causal level of the SSH. A **view** is the sensor reading at a **distinct** place, the omnidirectional image is a global reading of the surrounding at a certain place. Associating **view**s with omnidirectional images simplifies the data interpretation. Consider the following example. The robot stands in a distinct **place**. It takes an omnidirectional picture. It turns on the spot and then it takes a new picture. The new omnidirectional image will be the old one rotated by the same angle the robot rotated. The two pictures can be considered the same **view**, because of the rotational invariance. So, the robot will experience the same **view** before and after the **action** it took. This permits the robot to recognize that the **action** it took, was a **turn**, i.e. an action that leaves the agent at the same place. With a perspective camera the robot would have a totally different **view** after a turn on the spot. It would be really difficult to infer it is at the same **place**.

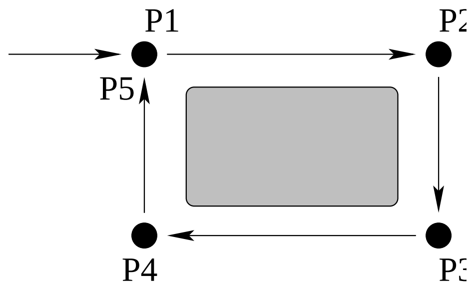


Fig. 5. The "exploring around the block" problem. The problem of recognizing the same place under different state labels.

The rotational invariance and the link between **views** and **actions** permit a straightforward solution to the problem of *exploring around the block*, i.e. the problem of recognising the same place under different state labels, see Figure 5. Here the robot is moving around the block following the arrows. When it reaches Place 5 from Place 4, it is very difficult to recognize it as the previously visited Place1, unless it is equipped with an omnidirectional camera and it makes use of the rotational invariance. Using the SSH terminology, it is easy to spot whether the current **view** is the same which has been experienced before and therefore to consider this view not as a different **place** but as the same **place** reached from a different direction.

The *distinctiveness measure* of the SSH permits to abstract **distinct** places from the continuous world. This is a function of the surrounding of the robot. An omnidirectional sensor permits the creation of a more effective distinctiveness measure that takes in to account all the feature of the world around the robot. As we will see in the Section VI, it is possible to identify in the omnidirectional frame sequence a set of events strictly related to the topological structure of the environment. These events correspond to the discontinuities of the distinctiveness measure [?]. The occurrence of such a discontinuity determines the transitions of state in the SSH.

[1] Azzurra Robot Team (ART) is the Italian team at Robocup Championship, the robot football competition.

[2] This mirror was designed and realized at Politecnico of Milano by Bonarini et al.

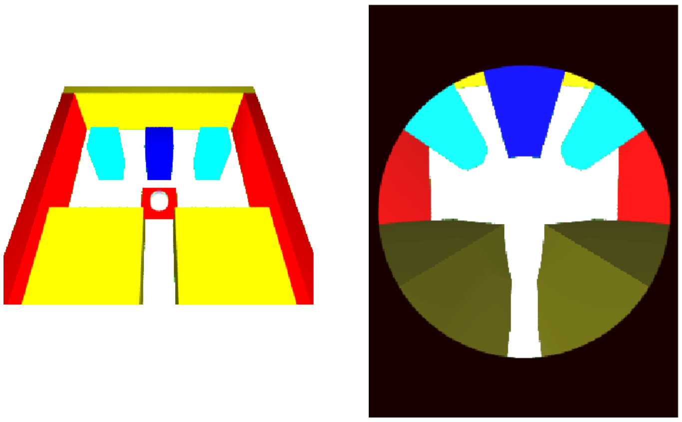
[3] In this figure a conical mirror is represented, but the properties which are illustrated apply to any kind of omnidirectional mirror.

[4] In the following, the bold font is used to indicate we are using the SSH meaning of the words.

V. Assumptions

In this work, we make use of some assumptions:

- The robot is moving in a indoor environment. This is a man-made environment like a building;
- The robot can only turn on the spot or move on a straight line. It cannot make more complex movements;
- The robot does not have direct access to the information on its movements;
- The lighting in the environment does not change during the motion of the robot;
- The objects present in the scene are static: they do not change their positions;
- The floor is almost flat and horizontal;
- The walls and the objects present in the scene have vertical edges and surfaces;
- The axis of the camera and the mirror are vertical;



(a) (b)

Fig. 6. (a) The perspective view of the virtual environment. The robot is the dark-gray square with the white sphere on top of it. (b) How the same scene is seen from the simulated omnidirectional sensor. Note that the body of the robot does not appear in the image.

VI. Features and Events Selection

We built a simulator to generate omnidirectional images of a virtual environment. The aim of the simulator was to permit us to carry out extensive preliminary test to extract clues about the features and the events that could be extracted from the omnidirectional image sequence. These features and events will be used to design the distinctiveness measure that abstracts distinct places from the continuous world. In this first part of the research, simulated images were preferred over the real images because the simulations provide an environment easily controllable and reconfigurable. Eventually, the vision system was tested both on simulated and real image sequences. The omnidirectional images were created using POV-Ray, a free software package for creating three-dimensional graphics (www.povray.org). The virtual environment is designed to present typical views of a man-made environment to the robot (i.e. corridors, doors, corners, objects, etc.), Figure 7.

Selected features are extracted from each omnidirectional image. When the robot moves the selected features appear to move in the sequence of omnidirectional images. The movement of the features originates the topological events we use within the SSH. These events happen at single points in the space, therefore they can be used to identify distinct points in the space. This is the key that permit us to extract from the continuous world a set of distinct places.

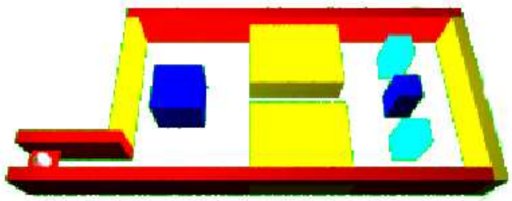


Fig. 7. The virtual environment

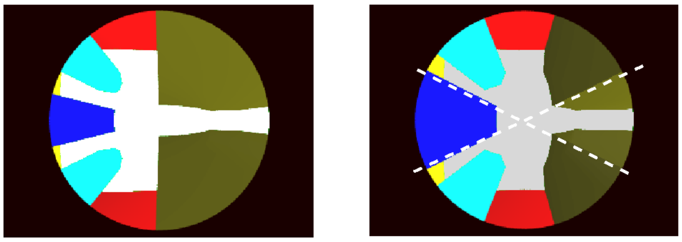
In the following we will discuss the features we decided to extract from the omnidirectional images and the topological events we identified in the frame sequence.

The features we extract from the pictures are the vertical edges present in the environment[?]. The vertical edges are features strictly binded to the objects present in the world and therefore easily recognizable by humans. Several authors selected non-intuitive features, instead, like brightness pattern or other image features only loosely related to the objects [?] [?]. We believe that for application like patrolling or remote surveillance the human readability is a must. Vertical edges present a double advantage: in a man-made environment like an office or a building, they are diffusely present [5] and they are easy to extract from the image. In fact, as mentioned in Paragraph III, the vertical edges are mapped into radial lines by the omnidirectional mirror. Therefore, they can be straightforwardly extracted with the use of a Hough transform [?] simplified by a opportune choice of the reference frame.

When the robot moves in the environment the vertical edges appear to move in the image. The movements of the vertical edges in the frame sequence generates the topological events. While the robot wanders several events happen: new objects come into view, other objects disappear from the image, the robot enters a door or a corridor, etc. Objects come into view either because the robot approaches an object that was too far away to be in the field of view or because the object is no longer *occluded* by another one. Objects enter in the field of view of the vision sensor more than six meters apart. This is a big distance from the sensor. Because we are interested in what happens in the

[5] Examples of vertical edges are doors, the sides of a cabinet, the legs of a chair, etc.

surrounding of the robot, we will focus only on the process of occlusion of the object's edges by other objects.



(a) (b)

Fig. 8. (a) Event 3: the robot enters a corridor, the two edges are 180 apart (b) Event 4: the robot sees two pairs of edges at 180.

The assumption of separate translations and rotations for the robot's movements permits to identify two sets of topological events occurring during the navigation of the robot.

During a translation the following events can happen:

1. a new edge exits from occlusion;
2. an edge disappear because occluded by another object;
3. two vertical edges are 180 apart in the vision sensor;
4. there are two pairs of vertical edges 180 apart.

Event 3 is particularly meaningful. In fact, it occurs when the robot enters a door or a corridor, i.e. a natural topological division of the space, see Figure 8. Each topological event causes a transition of state in our system, i.e. once one of these events occur a new **place** is abducted from the continuous space. The result is a segmentation of the space, see Figure 10

During a rotation there is no relative displacement from the robot and the objects. No edge appears or disappears. In other words, the image does not change, it is only rotated around its center. An invariance for rotation must be introduced in the distinctiveness measure.

VII. Experiments

We performed experiments in the simulated and in the real world. In these experiments, we tested the software for extracting the features and the events from the image sequences.

In the simulated experiments, the robot traveled through the virtual environment along the the path shown in Figure 9. The path is composed of two rotations and three translations. The vision system software is able to track the edges all along the path and to detect the topological events. The edges present in the picture are extracted with a Canny edge detector [?]. The tracking of the edges is done using the colour information present in the image. The vision software is able to recognize the turns and to retrieve the angle by which the robot turned. The output of the vision system is a division of the virtual environment into distinct places, in Figure 10 some of the segmenting lines encountered along the path are drawn. A new place is created every time a topological event is detected and the corresponding view is stored. In the end, we obtain a

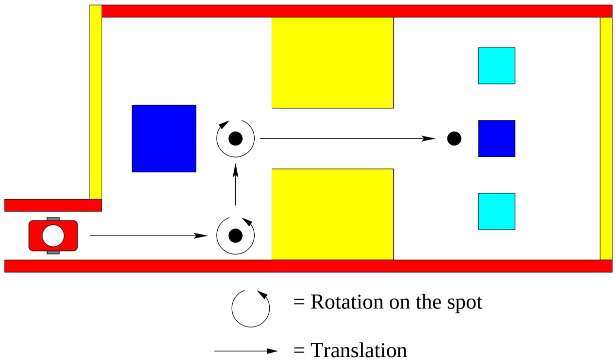


Fig. 9. The path of the robot through the virtual environment.

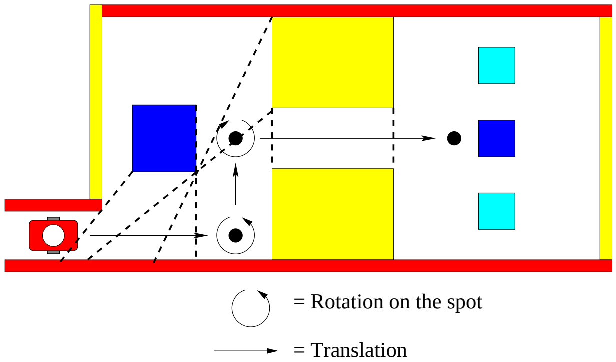


Fig. 10. The segmented path

topological map of places (associated to views) connected by actions: *travels* or *turns*.

In the experiments with the real robot, we encountered an implementational problem. Despite the vision system software worked properly in the simulations, the tracking of the vertical edges worked properly when the robot translated but it was not reliable when the robot turned on the spot. See Figure 11, for a picture acquired by the vision system of the real robot. This problem prevented the production of topological maps of paths containing **turns**. We discovered that the problem were caused by a reflection on

Fig. 11. An omnidirectional picture acquired by the real robot while it moves in the test environment.

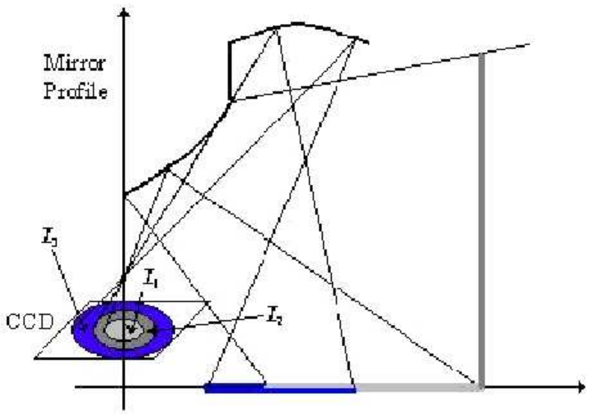


Fig. 12. The mirror profile of the newly designed multi-part mirror.

the Perspex cylinder. At the moment we are working to solve it.

VIII. Future work

At the same time, we designed a new mirror for the catadioptric sensor to overcome some limitation of the mirror currently used. This is a multi-part mirror where each segment is designed to view a specific area. The design of this mirror was inspired by the work of Marchese and Sorrenti [?]. To understand the mirror profile consider the sketch in Figure ??. The inner part of the mirror is designed to view objects from 60 cm around the robot up to six meters, without displaying the body of the robot (Part A in the Figure ??). This part produce the main part of the image. The remaining parts are more interesting. The middle ring (Part B) permits to view very distant objects and can be used for a better planning of the exploration movements, using the ideas about the catastrophe theory exposed in [?]. The external ring (Part C, sketched on the other side of the mirror for sake of clearity) displays at higher resolution (compared to the resolution attainable with the mirror we used in this paper) the area closer to the robot. This will be useful for the future design of more complex reactive control laws like *corridor following* and *wall following*. In fact, the current mirror is limited by the low resolution and the fact that the body of the robot hides the floor close to the robot. When we are writing, we just received this new mirror and we are carrying preliminary tests with it.

IX. Conclusion

In this paper, we showed that a catadioptric omnidirectional vision sensor is a good sensor for building a topological map using the Spatial Semantic Hierarchy. The *sensory level* has been successfully implemented (even if an unforeseen implementational problem mined the results on the real robot). The *control level* has been realized at a very primitive stage with very simple control laws and needs to be improved. We need to refine and to define clearly the *distinctiveness measure,* but the route is marked out. Our vision system showed that the egomotion of the robot can be estimated even without using active vision. The use of a new mirror designed for this application will improve the performances of the robot and produce new results.

Acknowledgments

We wish to thanks: Prof. Bob Fisher for his advice and support, Dr. John Hallam for his supervision to this project, Prof. Bonarini that gave us the omnidirectional mirror and the student of the ART-PD Robocup team who built the robot. This research has been partially supported by: the EC TMR Network SMART2, the Italian Ministry for the Education and Research (MURST), the Italian National Council of Research (CNR) and by the Parallel Computing Project of the Italian Energy Agency (ENEA).

References

[1] Andrea Bonarini. The body, the mind or the eye, first? In Manuela Veloso, Enrico Pagello, and Hiroaki Kitano, editors, *RoboCup99: Robot Soccer World Cup III*, volume 1856 pp. 210-221 of *LNCS*. Springer, 2000.

[2] J. Canny. A computational approach to edge detection. In *Transactions on Pattern Analysis and Machine Intelligence*, volume 8, pages 679–697. IEEE, 1986.

[3] P. Hough. Method for recognizing complex patterns. Technical report, US Patent 3069654, 1962.

[4] David Kortenkamp, Peter Bonasso, and Robin Murphy, editors. *Artificial Intelligence and Mobile Robotics.* AAAI Press/ MIT Press, 1998.

[5] B.J.A. Kröse, N. Vlassis, R. Bunschoten, and Y. Motomura. Feature selection for appareance-based robot localization. *Proceedings 2000 RWC Symposium*, 2000.

[6] Benjamin Kuipers. The spatial semantic hierarchy. *Artificial Intelligence*, 119:191–233, February 2000.

[7] Fabio Marchese and Domenico G. Sorrenti. Omni-directional vision with a multi-part mirror. *The fourth international workshop on robocup*, pages pp. 289–298, 2000.

[8] Carlos F. Marques and Pedro U. Lima. Vision-based self-localization for soccer robots. In *Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent robots and systems*, 2000.

[9] Francesco Nori and Enrico Pagello. Design of a multi-part omnidirectional mirror. Technical report, LADSEB CNR, Padua, July 2000.

[10] K. Pierce and B.J. Kuipers. Map learning with uninterpreted sensors and effectors. *Artificial Intelligence*, 92:169–227, 1997.

[11] Shoji Suzuki and Minoru Asada. An application of vision-based learing in robocup for a real robot with an omnidirectional vision system and the team description of osaka university "trackies". In Minoru Asada and Hiroaki Kitano, editors, *RoboCup98: Robot Soccer World Cup II*, volume 1604 pp. 316-325 of *LNCS*. Springer, 1999.

[12] Tomas Svoboda, Tomas Pajdla, and Vaclav Hlavac. Motion estimation using central panoramic cameras. *IEEE Conference on Intelligent Vehicles, Stuttgart, Germany*, October 1998.

[13] Niall Winters and Jose Santos-Victor. Mobile robot navigation using omni-directional vision. *3rd Irish Machine Vision and Image Processing Conference IMVIP99, Dublin, Ireland*, September 1999.

[14] M. Wright and G. Deacon. A catastrophe theory of planar orientation. *Int. Journal of Robotics Research*, 19(6):531–565, June 2000.

[15] Yasushi Yagi. Omni directional sensing and its applications. *IEICE TRANS. INF. & SYST.*, VOL. E82-D(NO. 3):pp. 568–579, MARCH 1999.

[16] Yasushi Yagi, Yoshimitsu Nishizawa, and Masahiko Yachida. Map-based navigation for a mobile robot with omnidirectionalimage sensor copis. *IEEE Transaction on Robotics and automation*, VOL. 11(NO. 5):pp. 634–648, October 1995.