DISTRIBUTED VISION SYSTEM FOR ROBOT LOCALISATION IN INDOOR ENVIRONMENT

E. Menegatti, G. Gatto, and E. Pagello[*]

Department of Information Engineering
Faculty of Engineering,
The University of Padua, Italy
{emg, gattogio, epv}@dei.unipd.it

Takashi Minato and Hiroshi Ishiguro

Department of Adaptive Machine Systems
Osaka University
Suita, Osaka, 565-0871 Japan
{minato, ishiguro}@ams.eng.osaka-u.ac.jp

ABSTRACT

Traditional image-based localisation methods do not work when the robot is moving in an environment whose appearance is changing in time. We propose an extension to the classical image-based localisation that uses a Distributed Vision System (DVS) and can work also in highly dynamic environments. The DVS is composed of omnidirectional cameras installed in the environment and can communicate with the robot. The localisation of the robot is achieved by comparing the current image grabbed by the robot with the images grabbed at the same time by the DVS. Finding the DVS's image most similar to the robot's image gives a topological localisation of the robot. In this paper, we analyze requirements and effectiveness of this approach and we present some preliminary experimental results obtained with the Distributed Vision System.

1. INTRODUCTION

Robot navigation in real world environments based on vision sensors is becoming more and more diffuse in robotics. In the past, several solutions have been proposed, but most of them can work only in static or almost static environments. Mobile robots should be able to work in dynamic environments in which many people are moving around or objects are displaced here and there. Most of the current localization systems assume the robot has a statical representation of the environment. The sensory inputs of the robot are processed and matched against the statical representation to find the robot position. In these approaches, a certain amount of noise or occlusion of the sensors can be tolerated [4, 13], but if the environment is changing too much in time (e.g. hundred people walking by, like in a metro station) the localization will not be successful.

Fig. 1. The humanoid robot Eveliee and the DVS installed in the environment

In the classical image-based localization approach, there is a setup stage in which the robot grabs several images at different locations in the environment and stores them in its memory. These images are called *reference images* and are annotated with the positions in which they were taken. At the running stage, when the robot moves around, to find its pose, it grabs an image and compares the current image with the image data-set stored in its memory. The reference image in memory most similar to the current image gives a topological localisation for the robot [1, 5, 17, 8, 18]. In previous works, we introduced an approach based on the Fourier signature of omnidirectional image [11] and showed that if combined with a Monte Carlo Localisation algorithm it could work reliably also in large environments [14]. However, all this works when the environment changes only slightly in time. If there are big changes the current image at a reference position will not match anymore the reference image stored in the memory of the robot. One possibility is to look at features that do not change even if the environment is crowded, like the ceiling [3]. However

[*]Also with Institute of Biomedical Engineering of the National Research Council (ISIB-CNR), Padova , Italy

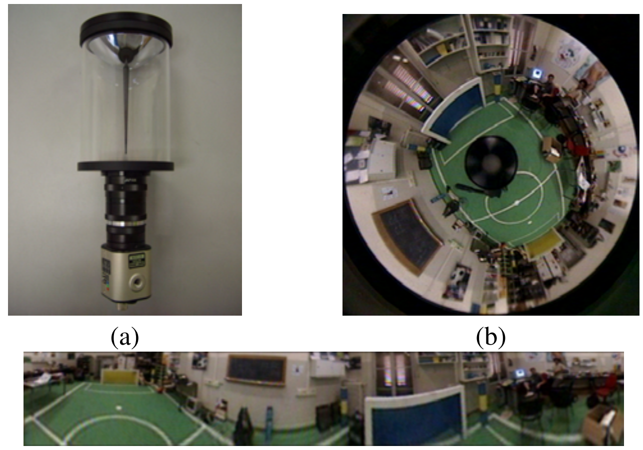
this is not always possible or do not solve the problem if there are lighting variations.

In this paper, we extend the image-based localisation approach to a dynamic environment. Our idea is: if a network of cameras is already installed in the environment (maybe for other purposes), this could be exploited as a Distributed Vision System (DVS). The robot's position can be estimated finding the camera whose image is most similar to the image grabbed by the robot. This approach will work both in static and in highly dynamic environment, because if the appearance of the environment changes, these changes are imaged both by the robot's camera and by the network's camera. This approach preserves all the advantages of the image-based localisation based on the Fourier signature proved in previous works: the possibility to have a hierchical coarse to fine localisation, the rotational invariance that enables to correctly estimate the position of the robot regardless of its orientation.

In this work we used the existing network of omnidirectional cameras available at the Intelligent Robotics Laboratory of Osaka University. The robot was equipped with an omnidirectional camera, as well. The basic assumption in this approach is that the image grabbed by the robot at any time and in any position in the environment is recognized as most similar to the image grabbed by the closest network's camera at the same time. It is easier to complain with this constraint, if one uses some of the technique we experienced in previous works: omnidirectional cameras [7][12], a measurement of similarity based on the Fourier signature [11], and grid computing for robotics applications [15].

The traditional approach to localize the robot in the environment with a network of camera is to detect the robot in the cameras images and determine its position thanks to previous camera calibration and possibly N-stereo geometry [10][9]. The problem with this approach is that visual features are not stable and it is not easy to correctly detect the robots in dynamic environments. Typical approaches like IR beacons and active or passive landmarks on the robots are not reliable in large environments. The problem is even more complex if we have several robot in the environment and we want to be able to identify them. In the future, the robot identification by visual feature might also be impossible for tiny or micro-robot, while for humanoid robots or android robots might be very hard to be able to distinguish them from humans walking by. Therefore, we developed this approach which does not depend on the appearance of the robot and do not require segmentation of visual features in the image.

It is important to note that the camera network is not dedicated to the robot localisation, but as is the case of the network used in this work, it has several other duties like surveillance, people tracking, people activity monitoring (in cooperation with floor pressure sensors and sound sensors) [6, 16]. Moreover, it should be noted that the robot is autonomous and can locate it-self using the system presented in [11, 14], the approach presented here can be integrated with this one to extend the robot abilities in case a DVS is available, but do not require the infrastructure to operate, like in [9].



(a) (b) (c)

Fig. 2. (a): An omnidirectional camera. (b): An example of omnidirectional image. (c): The panoramic cylinder obtained from the omnidirectional image (b).

2. DVS LOCALIZATION

The vision sensors being part of the DVS are not simple omnidirectional cameras. They can acquire the images, they can process them and can transmit over a LAN the result of the processing. Therefore, we call these vision systems Omnidirectional Vision Agents (OVAs) to stress their ability to process the image and to communicate with other agents.

The localisation system presented in this work is composed of several static OVAs (Omnidirectional Vision Agents) comparing their images to find which one is more similar to the images of the mobile OVA.

The localisation system must work in real-time, so one of the first issues is to have a system able to provide fast transmission and processing of the images to enable a real-time comparison of the images. Therefore, we need to design the system which requirements are: (i) to minimize the amount of data needed to describe the images; (ii) to maximize the speed of calculating the similarity among the images; (iii) to have a complete description of the scene to reliably assess the similarity of the images.

The first variable to be optimized is the number of cameras. By using omnidirectional cameras, we keep this number to a minimum, see Fig. 2a). In fact, every omnidirectional camera has a 360° field of view and images the whole scene surrounding it, Fig. 2b). This means a complete description of the scene and less cameras needed to cover the whole environment.

To reduce the amount of data transmitted and to ensure a fast calculation of the similarity between the images, the OVA does not send the whole image over the net, but sends a compact representation, called the *Fourier signature* of the image.

2.1. The Fourier signature and its properties

We showed in previous works the Fourier signature is a complete and compact representation of an omnidirectional image with many interesting proprieties [11]. To obtain the Fourier signature, the omnidirectional image is 'unwarped' into a panoramic cylinder, Fig. 2 (c), and then the Fourier transform of every row of the panoramic cylinder is calculated. The Fourier signature of every image is represented by the first 15 Fourier coefficients of the magnitude and of the phase associated to every row of the panoramic cylinder. The robot's position can be extracted using the magnitude coefficients of the Fourier signature, while the robot's orientation can be extracted by the phase coefficients. For more details please refer to [11].

One of the main advantages of the use of the Fourier signature is its intrinsic rotational invariance, which enables to match the correct OVA image regardless of the orientation of the robot (at the same time the orientation of the robot can be calculated using the phase values). In addition to the rotation invariance, another fundamental property of the Fourier signature essential in this application is what we called the *perspective invariance*. Consider the case sketched in Fig. 3, where a person is standing between the OVA and the robot. In this situation, the omndirectional cameras of the OVA and of the robot see the approximately the same omnidirectional image. The person standing by seen occluding the right part of the background by the OVA, while it is seen occluding the left part of the background by the robot, Fig. 9 (b) and (c). Nevertheless, the Fourier signatures of the two images do not differ greatly (we tested the robustness to occlusion in [19]). In fact, calculating the Fourier transform only along the rows of the panoramic cylinder results in an invariance to the horizontal distribution of brightness pattern. Therefore, it is not important the horizontal position of the occlusion in the images, but the fact that the occlusion is there or not.

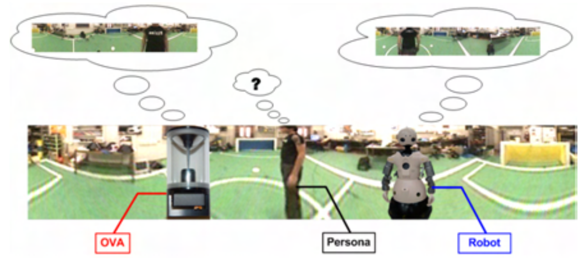


Fig. 3. The concept of the *perspective invariance*.

2.2. Localisation in Dynamic Environment

The robot localisation process starts when the robot send a localisation request over the wireless LAN in broadcast to the OVAs in reach of the wireless card. The localisation request contains the Fourier signature of the image grabbed by the robot. Every VA that receives the localisation request, grabs an image and calculates the Fourier signature and compares this with the Fourier signature sent by the robot using the similarity function defined in [11]. The VA's image more similar to the robot's image gives the topological localisation of the robot.

To test the effectiveness of the approach we compare the new approach with the classical image-based localisation approach. As we said, in the classical image-based localisation approach there is a set-up stage in which the robot stores in its memory the dataset of the reference images. If at the running phase the environment changes (like in Fig. 5 where many people are passing by, or like in Fig. 7, where a big variation in the lighting of the environment occured), the current image grabbed by the robot cannot longer match the correct reference image stored in the robot's memory. In the experimental sections we will show that if the current robot's image is compared with the DVS current images, the correct position can be calculated.

2.3. Requirements

The requirements for a reliable localisation are listed below. Experimentally we found that some of them heve to be strictly met and others can be met only loosely. These requirements are related mostly to the current implementation of this idea and dependent on the technique adopted to asses the similarity between the omnidirectional images (i.e. the similarity Function defined on the Fourier signature). We are working on finding new similarity function

Fig. 4. Image-based localisation with a classical approach: with and without people moving around.

that will enable to relax these requirements, in order to be able to include different type of cameras as well, e.g. perspective cameras.

1. the robot camera should be as close as possible to the OVAs; this ensures the images grabbed by the robot and by the closest OVA are similar.
2. there shouldn't be walls or very wide objects between the OVA and the robot. If there is a wall limiting the view of the robot and the OVA is over the wall, like in Fig. 1, even if the robot and the OVA are really close their views will be really different.
3. the robot's camera and the OVAs' cameras should be of the same kind. Different cameras or different omnidirectional mirror can results in different omnidirectional images with very different Fourier signatures.
4. the robot's camera and the OVAs' cameras should be located approximately at the same height. Too different heights can cause different perspective distortions and result in erroneous localisation.

These requirements are very strict for a pratical implementation, indeed. However, in the following we show the approach is working also when using a DVS not built on purpose for this application. The DVS used is a pre-existing network of cameras used for other tasks. This network does not fulfill all the above requirements. In particular, the OVAs are fixed over panel walls dividing the room (requirement 2), so sometimes the robot sees the wall, but the camera do not see it.

3. EXPERIMENTS

In this section we do not present the whole localisation with the Monte Carlo algorithm, but just some tests in which a static image-based localisation would have failed. The image Fig. 6 a) is the one grabbed by the robot, the image Fig. 6 b) is the one grabbed by the closest OVA. Thanks to the rotational invariance and the other property of the Fourier signature they are recognised as the same image from the localisation system (both from the classical one and the new one using the DVS).

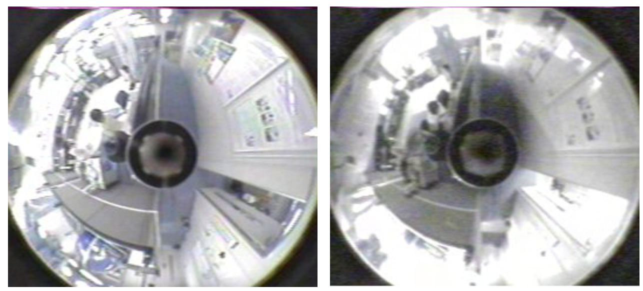
3.1. Comparison between classical and DVS approach

If the appearance of the environment changes after the set-up phase, like in Fig. 7 where the lighting conditions are

Fig. 5. Image-based localisation with the proposed DVS approach with people walking by.

(a) (b)

Fig. 6. An example of an omnidirectional image grabbed by the robot (a) and by the closest OVA (b).



(a) (b)

Fig. 7. (a): Two omnidirectional images grabbed by the robot in the same position: (a) at the set-up stage with the lights on; (b) at the running phase with the lights off. Note the auto-gain of the camera is turned on, so the average brightness is the same, but the images are different.

changed and the lights are switched off. The results given by the two methods are no longer the same. The classical method gives an erroneous localisation, Fig. 8a). Note the correct location achieve a similarity value rather low, it is the sixth value. While the proposed DVS method correctly localises the robot, Fig. 8b). This is because if the lights go off, also the images grabbed by DVS will be dimmer, i.e. robot and OVAs see the same variations.

3.2. Perspective Invariance

To test the robustness of the proposed approach to the perspective invariance, we recreated a situation similar to the one described in Section 2.1: a person is standing between the OVAs and the robot. For this experiment we have 6 OVAs like in Fig. 9a). A person is standing between OVA 4 and the robot that is near OVA 4.

The images grabbed by OVA 4 and by the robot are very different (the person is seen against different portions of the background). They are shown in Fig. 9 (b) and (c). Nevertheless, thanks to the perspective invariance introduced in Section 2.1 the two images are correctly matched as show the similarity values plotted in Fig. 9 b)

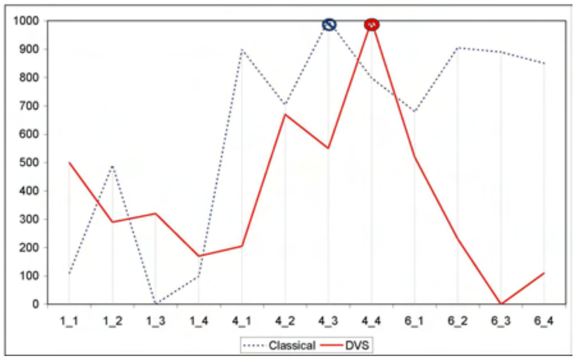
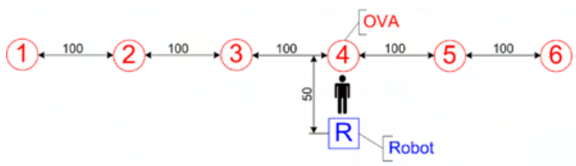
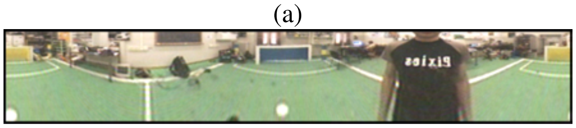


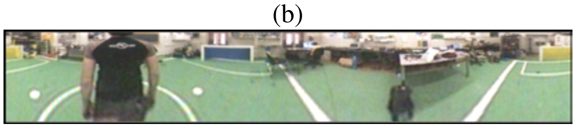
Fig. 8. Plot of the current robot image similarity against the DVS images (solid line) and the reference images in the robot's memory (dashed line).



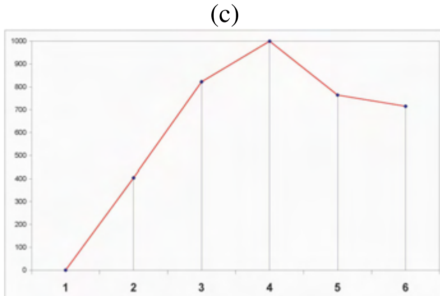
(a)



(b)



(c)



(d)

Fig. 9. (a)A sketch of the relative positions of the OVAs and of the robot; (b) The same environment observed by the robot and (c) by the closest OVA when a person is standing in between; (d) The plot of the similarity values calculated for the robot's image.

4. STRENGTHENING THE SYSTEM

At the moment we are working to make the system more flexible, in order to be able to relax the constraints and the requirements introduced in Section 2.3. The ultimate reason for the necessity of so strict requirements is the image similarity function that discard a lot of information to focus only on the brightness pattern described by the Fourier signatures. This might results in reference images very different from the input image, but with a similar Fourier signature. We are experimenting new similarity measures that could take into account the information discarded by the Fourier signature. Our aim is to be able to cluster the DVS images depending on their appearance, to determine to which cluster the robot's image is more similar and then to use the Fourier signature similarity only in the selected image cluster to exploit the properties of the Fourier signature localisation. We developed an image similarity function based on colors which has the same two fundamental properties of the Fourier signature: (i) the rotational invariance and (ii) the perspective invariance. We take into account the H (hue) and S (saturation) components of the HSV colour space. This similarity was inspired by [2]. Fig. 10 shows that the HS-diagrams are similar for OVAs nearby, while the HS-diagrams are very different for OVAs far away.

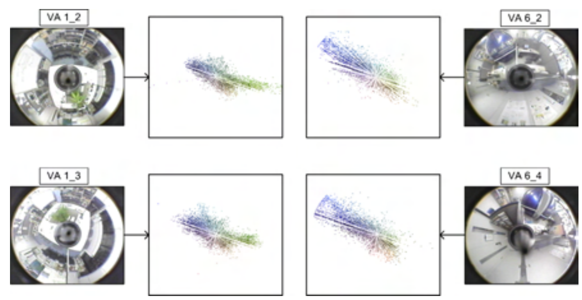


Fig. 10. A comparison of HS-diagrams for OVAs close-by and far away.

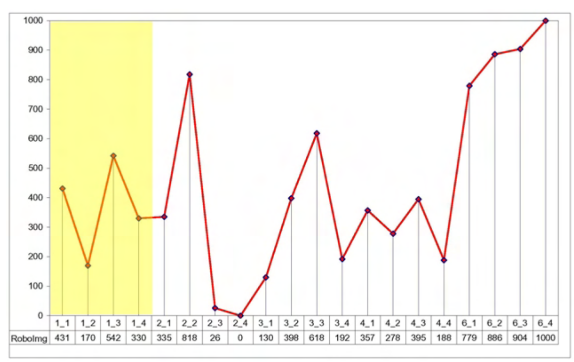


Fig. 11. Plot of similarity values showing that using the HS-clustering a correct localisation can be achieved even if the absolute maximum of the similarity is not at the correct location.

The plot of the similarity values of the robot image against all DVS images in Fig. 11 shows that a correct localisation can be achieved also when the requirements of Section 2.3 are not met. In this last experiments, we introduced OVAs with a significant difference in height from the robot's camera, which did not satisfied Requirement 4. The plot in Fig. 11 shows an incorrect robot localisation because the highest value is at 6_4, while the robot is located at 1_3. However, if the HS-clustering algorithm is used, the system finds that the robot image belongs to the cluster of cameras (1_1, 1_2, 1_3, 1_4), the area highlighted in yellow in Fig. 11. So, if the Fourier signature similarity is calculated only within the correct cluster (the yellow one), the maximum similarity in this cluster is obtained at the correct location 1_3.

5. CONCLUSIONS

In this paper we proposed an extension for a previously proposed image-based localisation approach. The extension is based on a Distributed Vision System (DVS). To find its location, the mobile robot compares its current image with the images grabbed at the same time by the DVS. The calculation of the similarity is based on the Fourier signature we introduced in previous works. We identified some constraints the system have to meet in order to give correct results. The system has been tested using a humanoid robot and a pre-existing camera network. The experiments showed the system can cope with dynamic environments where the classical image-based approach fails. At the moment we are working to relax the requirements we pinpointed in the text in order to have a more flexible system. In the end, we hinted the direction we are taking in order to be able to relax those constraints.

6. REFERENCES

[1] H. Aihara, N. Iwasa, N. Yokoya, and H. Takemura. Memory-based self-localisation using omnidirectional images. In A. K. Jain, S. Venkatesh, and B. C. Lovell, editors, *Proc. of the 14th International Conference on Pattern Recognition*, volume vol. I, pages 1799–1803, 1998.

[2] D. Balthasar. Color matching by using tuple matching. *International Conference on Image Analysis and Processing*, 1(12):402–407, 9 2003.

[3] F. Dellaert, W. Burgard, D. Fox, and S. Thrun. Using the condensation algorithm for robust, vision-based mobile robot localization. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'99)*, June 1999.

[4] D. Fox, W. Burgard, and S. Thrun. Markov localization for mobile robots in dynamic environments. *Journal of Artificial Intelligence Research*, 11, 1999.

[5] J. Gaspar, N. Winters, and J. Santos-Victor. Vision-based navigation and environmental representations with an omnidirectional camera. *IEEE Transaction on Robotics and Automation*, Vol 16(number 6), December 2000.

[6] T. Ikeda, T. Ishida, and H. Ishiguro. Framework of distributed audition. In *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN)*, page 1B4, 2004.

[7] H. Ishiguro. Development of low-cost compact omnidirectional vision sensors. In R. Benosman and S. Kang, editors, *Panoramic Vision*, chapter 3, pages 23–38. Springer, 2001.

[8] B. Kröse, N. Vlassis, R. Bunschoten, and Y. Motomura. A probabilistic model for appareance-based robot localization. *Image and Vision Computing*, vol. 19(6):pp. 381–391, April 2001.

[9] E. Kruse and F. Wahl. Camera-based monitoring system for mobile robot guidance. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS98), pages=pp. 1248-1249, year = 1998*.

[10] T. Matsuyama, S. Hiura, T.Wada, K.Murase, and A.Yoshioka. Dynamic memory: Architecture for real time integration of visual perception, camera action, and network communication. In W. Burgard, T. Christaller, and A. B. Cremers, editors, *Proc. of Computer Vision and Pattern Recognition Conference*, pages pp. 728–735, June 2000.

[11] E. Menegatti, T. Maeda, and H. Ishiguro. Image-based memory for robot navigation using properties of the omnidirectional images. *Robotics and Autonomous Systems, Elsevier*, 47(4):pp. 251–267, July 2004.

[12] E. Menegatti, F. Nori, E. Pagello, C. Pellizzari, and D. Spagnoli. Designing an omnidirectional vision system for a goalkeeper robot. In A. Birk, S. Coradeschi, and S. Tadokoro, editors, *RoboCup-2001: Robot Soccer World Cup V.*, L. N. on A. I, pages 78–87. Springer, 2002.

[13] E. Menegatti, A. Pretto, and E. Pagello. Testing omnidirectional vision-based Monte-Carlo localization under occlusion. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS04)*, pages pp. 2487–2494, September 2004.

[14] E. Menegatti, M. Zoccarato, E. Pagello, and H. Ishiguro. Image-based Monte-Carlo localisation with omnidirectional images. *Robotics and Autonomous Systems, Elsevier*, 48(1):17–30, August 2004.

[15] E. Mumolo, E. Menegatti, and E. Pagello. Omnidirectional image based localisation using an xml-based language for distributed computing. In *Proc. of 8th International Conference on Intelligent Autonomous Systems (IAS-8)*, pages 943–951, Amsterdam - The Netherlands, Month 2004.

[16] T. Murakita, T. Ikeda, and H. Ishiguro. Human tracking using floor sensors based on the markov chain monte carlo method. In *Proceedings of the Seventeenth International Conference on Pattern Recognition (ICPR)*, pages 917–920, 2004.

[17] T. Wilhelm, H.-J. Böhme, and H.-M. Gross. A multi-modal system for tracking and analyzing faces on a mobile robot. *Robotics and Autonomous Systems*, 48:pp. 31–40, August 2004.

[18] J. Wolf, W. Burgard, and H. Burkhardt. Robust vision-based localization by combining an image retrieval system with monte carlo localization. *IEEE Transactions on Robotics*, 21(2):208–216, 2005.

[19] M. Zoccarato. Localizzazione di monte carlo per robot mobili dotati di visione omnidirezionale. Master's thesis, Univ. of Padua, 2002–2003. (in Italian only).