

Omnidirectional stereo systems for robot navigation

Giovanni Adorni(*)
Monica Mordonini(+)

Stefano Cagnoni(+)
Antonio Sgorbissa(*)

(+) Dept. of Computer Engineering
University of Parma
Parma, Italy 43100

(*) DIST
University of Genoa
Genoa, Italy 16145

Abstract

This paper discusses how stereo vision achieved through the use of omnidirectional sensors can help mobile robot navigation providing advantages, in terms of both versatility and performance, with respect to the classical stereo system based on two horizontally-displaced traditional cameras.

The paper also describes an automatic calibration strategy for catadioptric omnidirectional sensors and results obtained using a stereo obstacle detection algorithm devised within a general framework in which, with some limitations, many existing algorithm designed for traditional cameras can be adapted for use with omnidirectional sensors.

1. Introduction

The need for robotic sensory systems that provide a global description of the surrounding environment is increasing. In mobile robotics applications, autonomous robots are required to react to visual stimuli that may come from any direction at any moment of their activity, and to plan their behaviour accordingly. This has stimulated growing interest in omnidirectional vision systems [1]. Such systems provide the widest possible field of view and obviate the need for active cameras that require complex control strategies, at the cost of reduced resolution with respect to traditional cameras, that distribute a smaller field of view on the same sensor surface.

Recent robotics and surveillance applications in which omnidirectional sensors have been used effectively, as the only vision sensor or jointly with other higher-resolution non-omnidirectional ones, are described in [2, 3, 4].

Applications of mobile robotics in which robots rely on vision for safe and efficient navigation share a set of features and requirements, often conflicting with one another, that strongly influence application design criteria. Among these:

- robots are immersed in a dynamic environment that may change quite rapidly, within and beyond their field of action;

- robots require high-resolution vision for accurate operation within their field of action;
- robots need wide-angle vision to be aware of what happens beyond their field of action and to react/plan accordingly.

Regarding the typical environment where virtually all indoor robotics and most outdoor robotics take place, further considerations can be made about the natural partial structuration of the space in which mobile robots operate. Such a space is usually inferiorly delimited by the plane (floor/ground) on which robots move and extends vertically up to where robots can see or physically reach. The floor can therefore be assigned the role of reference plane in the main tasks in which mobile robots are routinely engaged during navigation, namely self-localization, obstacle detection, free-space computation. We could therefore call it a 2D augmented environment, to underline that the two dimensions along which the floor extends are privileged with respect to the third dimension. Even within these limitations, robots actually operate in a 3D environment and their operation can take advantage of 3D information. Stereo vision is therefore appealing to several navigation tasks.

However, traditional stereo vision setups, made up of two traditional cameras displaced horizontally, hardly satisfy the above-mentioned requirements of autonomous robotics applications. The use of omnidirectional sensors, besides providing the robot with obvious advantages in terms of self-localization capabilities, can be extremely useful also to extract 3D information from the environment using stereo algorithms.

In section 2 we introduce a sensor model, based on the joint use of an omnidirectional sensor and a traditional one, with which powerful stereo algorithms can be implemented. We then briefly compare such a model with traditional and fully-omnidirectional stereo setups. In section 3 we propose a framework within which a particular class of algorithms for omnidirectional sensors can be easily developed, as an extension of traditional stereo algorithms, with almost no

extra overhead. Such a class of algorithms that are applicable to 2D augmented environments, which includes many if not most real-world applications, can be termed the quasi-3D (q3D) class. More precisely, it comprises algorithms that can exploit the presence, in the environment, of a reference plane for which a transform (the Inverse Perspective Transform) exists, which allows for the recovery of visual information through a remapping operation. A fast and simple auto-calibration process that allows for such a mapping is described in section 4. In section 5, as an example, we eventually describe the basics of an efficient obstacle detection algorithm developed within this framework.

2. Hybrid and fully-omnidirectional stereo vision sensors

Using traditional stereo systems, typically made up of two traditional cameras aligned and displaced along the horizontal axis, has several drawbacks in mobile robot applications. Among them:

- the constraints imposed by the configuration of the two traditional cameras needed to obtain sufficient disparity often conflict with the general requirements of the applications for which the stereo system is used;
- the resulting field of view of the stereo system is much smaller than the, already limited, field of view of each of the two cameras.

The first drawback mainly affects robot design, since it requires that a front and a rear side of the robot be clearly defined. This can be a severe limitation when holonomous robots are used. With a traditional stereo setup, any reconfiguration of the (strongly asymmetric) vision system requires that both cameras be repositioned and might possibly call for structural modifications.

The second drawback is particularly relevant in dynamic environments. If one considers that a robot movement should be ideally exclusively finalized to performing the task of interest, it is immediately evident how penalizing it is for the robot having to move itself just to second its own perceptual needs.

Using omnidirectional sensors is beneficial with regard to both problems. Here, we consider two models, a hybrid omnidirectional/pin-hole system and a fully-omnidirectional one.

In particular, it is clear that a symmetric coaxial fully-omnidirectional model as the one briefly discussed in section 2.1 can solve both problems. However, the solution comes at the cost of a lower resolution in the far field and of the loss of horizontal disparity between the two views, which may be also unacceptable in some applications.

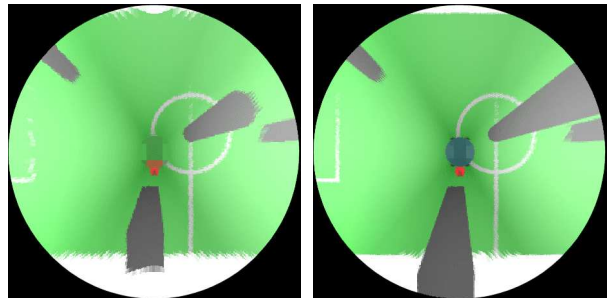


Figure 1: A fully-omnidirectional sensor model (above) and the Inverse Perspective Transform (see section 3) images of a simulated RoboCup field with four robots and a ball obtained with such a mirror configuration (upper sensor below on the left, lower one below on the right).

A way to obtain stereo images, providing the robot with both low-resolution omnidirectional vision in the far field and high-resolution vision in the near field while keeping the field of view as wide as possible, is to use a sensor made up of both an omnidirectional camera and a traditional one.

In the following, after showing the results of a simulation of a fully-omnidirectional system to provide a feeling of how images acquired by such systems may look like, we describe in details HOPS (Hybrid Omnidirectional/Pin-hole System), a stereo model that tries to achieve a good trade-off, with particular attention to mobile robot applications, between the features provided by omnidirectional and traditional systems.

2.1. Fully omnidirectional model

A fully-omnidirectional stereo model uses two omnidirectional sensors for stereo-disparity computation. In figure 1 we show preliminary results of a simulated vision systems made up of two catadioptric omnidirectional sensors. We have taken into consideration a configuration in which the vision sensors are placed one above the other, and share a common axis Z (figure 1, above on the right) perpendicular to the reference plane.

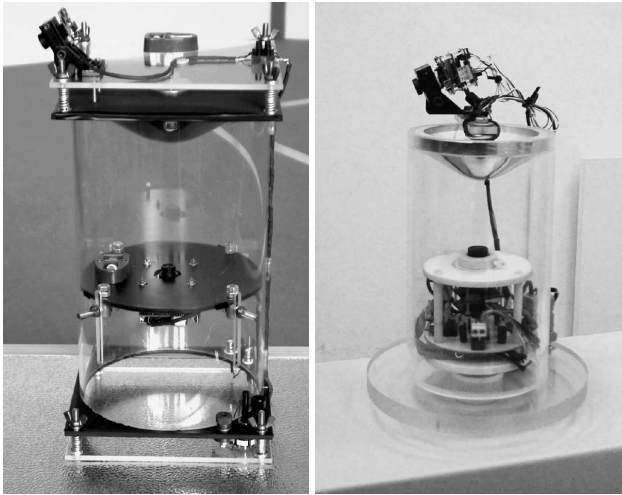


Figure 2: The two hybrid sensor prototypes: HOPS1 and HOPS2.

The main drawback of such a coaxial configuration is that it provides no lateral stereo disparity (see section 5), which make obstacles recognizable only exploiting vertical stereo disparity. On the other hand, dealing with a stereo sensor having two sensors with parallel axes is more complicated, both in terms of construction, size and calibration.

2.2. Hybrid omnidirectional/pin-hole model

HOPS (of which two prototypes are shown in figure 2) is a hybrid vision sensor that integrates omnidirectional vision with traditional pin-hole vision, to overcome the limitations of the two approaches. If a certain height is needed by the traditional camera to achieve a reasonable field of view, the top of the omnidirectional sensor may provide a base for the traditional CCD-camera based sensor that can lean on it, as shown in figure 2. In the prototype shown in figure 2a the traditional camera is fixed and looks down with a tilt angle of about 60° with respect to the ground plane and has a field of view of about 80° . To obtain both horizontal and vertical disparity between the two images, it is positioned off the center of the device. The 'blind sector' caused by the upper camera cable on the lower sensor is placed at an angle of 180° with respect to a conventional 'front view', in order to relegate it to the back of the device. If a lower point of view is acceptable for the traditional camera, it can also be placed below the omnidirectional sensor, provided it is out of the field of view of the latter. The top of the device is easily accessible, allowing for easy substitution of the catadioptric mirror. Consequently, also the camera holder on which the upwards-pointing camera is placed can be moved upwards or downwards, to adjust its distance from the mirror. In the prototype in figure 2b, the traditional camera is positioned laterally above the omnidirectional sensor on a holder that

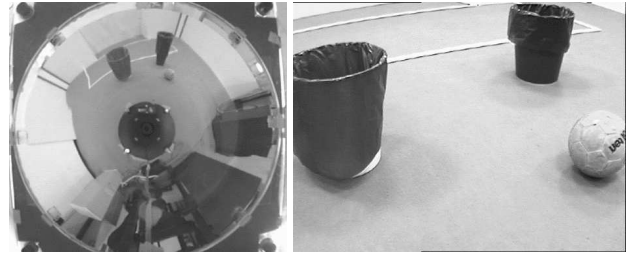


Figure 3: Example of images that can be acquired through the omnidirectional sensor (left) and through the CCD camera (right) of the HOPS1 prototype.

can be manually rotated.

An example of the images that can be acquired through the two sensors of the first prototype is provided in figure 3.

The aims with which HOPS was designed are accuracy, efficiency and versatility. The joint use of a standard CCD camera and of an omnidirectional sensor provides HOPS with different and complementary features: while the CCD camera can be used to acquire detailed information about a limited region of interest, the omnidirectional sensor provides wide-range, but less detailed, information about the surroundings of the system. HOPS, therefore, suits several kinds of applications as, for example, self-localization or obstacle detection, and makes it possible to implement peripheral/foveal active vision strategies: the wide-range sensor is used to acquire a rough representation of a large area around the system and to localize the objects or areas of interest, while the traditional camera is used to enhance the resolution with which these areas are then analysed. The different features of the two sensors can be exploited in both a stand-alone way as well as in a combined use. In particular, as discussed in section 5, HOPS can be used as a stereo sensor to extract three-dimensional information about the scene that is being observed.

3. General framework for stereo algorithm development

Images acquired by the cameras on-board the robots are affected by two kinds of distortions: perspective effects and deformations that derive from the shape of the lens through which the scene is observed. Given an arbitrarily chosen reference plane (typically, the floor/ground on which robots move), it is possible to find a function $P_{x,y} = \mathcal{C}(I_{i,j})$ that maps each pixel in the image $I_{i,j}$ onto the corresponding point $P_{x,y}$ of a new image P (with coordinates x, y) that represents a bird's view of the reference plane. Limiting one's interest to the reference plane, it is possible to reason on the scene observing it with no distortions. The most appealing feature, in this case, is that a direct correspondence

between distances on the reconstructed image and in the real world can be obtained, which is a fundamental requirement for geometrical reasoning. This transformation is often referred to as Inverse Perspective Transform (IPT) [5, 6, 7], since perspective-effect removal is the most common aim with which it is performed, even if it actually represents only a part of the problem for which it provides a solution.

If all parameters related to the geometry of the acquisition systems and to the distortions introduced by the camera were known, the derivation of \mathcal{C} could be straightforward. However, this is not always the case, most often because of the lack of an exact model of camera distortion. However, it is often possible to effectively (and efficiently) derive \mathcal{C} empirically using proper calibration algorithms, as shown in next section.

The IPT plays an important role in several robotics applications in which finding a relevant reference plane is easy. This is true for most indoor Mobile Service Robotics applications (such as surveillance of banks and warehouses, transportation of goods, escort for people at exhibitions and museums, etc.), since most objects which the robot observes and with which it interacts lie in fact on the same plane surface of the floor on which the robot is moving. Since our system has been mainly tested within the RoboCup¹ environment, in the following we will take it as a case study. In RoboCup everything lies on the playing field and hardly raise significantly above, as happens, for example, with the ball. Therefore, the playing field can be taken as a natural reference plane.

In the rest of the paper we will show how a general empirical IPT mapping can be applied, even more effectively, also to catadioptric omnidirectional sensors. The intrinsic distortion of such sensors, especially with respect to the typical images with which humans are used to dealing, makes direct image interpretation difficult, since a different reference system (polar coordinates) is implicitly 'embedded' in the images thus produced. However, their circular symmetry allows for a simplification of the IPT computation.

Exploiting this feature in implementing the IPT for catadioptric omnidirectional sensors, we have devised an efficient automatic calibration algorithm that will be described in the next section.

4. Omnidirectional sensor calibration

In computing \mathcal{C}_o , the generalization of the IPT for a catadioptric omnidirectional sensor, the problem is complicated by the non-planar profile of the mirror; on the other hand, the circular symmetry of the device provides the opportunity of dramatically simplifying such a procedure.

If the reflecting surface were perfectly manufactured, it would be sufficient to compute just the restriction of \mathcal{C}_o

along one radius of the mirror projection on the image plane to compute the whole function. However, possible manufacturing flaws may affect both the mirror shape and the smoothness of its surface. In addition to singularities that do not affect sensor symmetry and can be included in the radial model of the mirror (caused, for example, by the joint between two differently shaped surfaces required by the specifications for a particular application, as in [8]), a few other minor isolated flaws can be found scattered over the surface. Similar considerations can be made regarding the lens through which the image reflected on the mirror is captured by the camera.

To account for all sorts of distortions an empirical derivation of \mathcal{C}_o based on an appropriate sampling of the function in the image space can be made. Choosing such a procedure to compute \mathcal{C}_o permits to include also the lens model into the mapping function.

The basic principle by which \mathcal{C}_o can be derived empirically is to consider a set of equally-spaced radii, along each of which values of \mathcal{C}_o are computed for a set of uniformly-sampled points for which the relative position with respect to the sensor is known exactly. This produces a polar grid of points for which the values of \mathcal{C}_o are known.

To compute the function for a generic point P located anywhere in the field of view of the sensor, a bi-linear interpolation is made between the four points, belonging to a uniformly-sampled polar grid, among which P is located. This makes reconstruction accuracy better in proximity of the robot, as the actual area of the cells used for interpolation increases with radial distance while, correspondingly, image resolution decreases. The number of data-points (interpolation nodes) needed to achieve sufficient accuracy depends mainly on the mirror profile (the smoothest the profile, the fewest the points) and on the mirror surface quality (the fewest the flaws, the fewest the points).

This calibration process can be automated, especially in the presence of well manufactured mirrors, by automatically detecting relevant points. To do so, a simple pattern consisting of a white stripe with a set of aligned black squares superimposed on it can be used, as shown in figure 4.

The reference data-points, to be used as nodes for the grid, are extracted by automatically detecting the squares in a set of one or more images grabbed turning the robot around the vertical axis of the sensor. Doing so the reference pattern is reflected by different mirror portions in each image.

Using different shapes instead of squares, e.g., circles or ellipses, is obviously possible: using appropriate ellipses in points located far from the center of the mirror could even be advantageous, because they could be seen approximately as circles in the grabbed images, simplifying their recognition. In any case, if distances between the shapes form-

¹ visit <http://www.robocup.org> for more information.

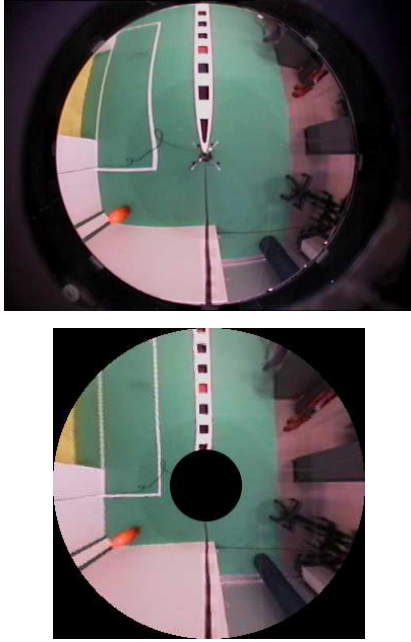


Figure 4: The pattern used for calibrating a catadioptric omnidirectional sensor (above). The fourth square from the center has a different color, to act as a landmark in automatically computing distances; below it the IPT image obtained after calibration is shown. The black circle hides the expansion of the area, roughly corresponding to the robot footprint, whose reflection is removed in the original image by providing the mirror with a discontinuity in its center.

ing the pattern are known exactly, the only requirement is that one of the shapes, at known distance, be distinguishable (e.g., by its color) from the others. The shape should be possibly located within the highest-resolution area of the sensor. This permits to use the reference shape as a landmark to automatically measure the distance from the camera of every shape on the reference plane. This also removes the need to accurately position the robot at a predefined distance from the pattern, which could be a further source of calibration errors.

Operatively, in the first step of the automatic calibration process, the white stripe, as well as the center of every reference shape, are easily detected. These reference points are inserted into the set of samples on which interpolation is then performed. The actual position of such points can be simply derived from the knowledge of the relative position of the square pattern to which it belongs with respect to the reference differently-colored shape. The process can be repeated for different headings of the robot, simply turning the robot around its central symmetry axis.

In the second step, interpolation is performed to compute the function \mathcal{C}_o from the point set extracted as described.

A look-up table that associates each pair of coordinates in the IPT-transformed image to a pair of coordinates in the original image can thus be computed.

This calibration process is fast, can be completely automated and provides good results, as shown in figure 4.

5. Experiments with an IPT-based obstacle detection algorithm for omnidirectional sensors

As an example of algorithm porting from traditional stereo systems to omnidirectional ones using the generalized IPT, we report some sample results, obtained in a robot soccer environment, of a stereo algorithm for obstacle detection developed for traditional stereo systems [5] and adapted for use with HOPS. The algorithm is described in details elsewhere [9]: here we mainly aim at showing its potentials and evidentiating the role played by the generalized IPT.

Besides removing the distortion introduced by the omnidirectional sensor using the IPT, the algorithm exploits the intrinsic limitation of the IPT to be able to provide undistorted views only of the objects that lie on one reference plane. Everything that is above the plane is distorted differently as a function of its height and of the point of view from which it is observed. Therefore, two IPT-transformed images of the same scene will differ only in those regions that represent obstacles, i.e., any object located above the reference plane. In mobile robotics applications, the reference plane is chosen to be the floor on which the robots are moving.

Given two images of the same spatial region that includes the floor on which a robot is moving, the obstacle detection algorithm can be roughly summarised as follows:

1. compute the IPT of both images with respect to the plane identified by the floor;
2. apply an edge extraction algorithm to the IPT-transformed images;
3. skeletonize and binarize the contours using a ridge-following algorithm;
4. compute the difference between the two images obtained in the previous step.

When the chromatic features of the two images obtained from the two sensors are virtually identical the steps 2 and 3 of the algorithm can also be substituted by a thresholding algorithm by which objects that clearly stand out with respect to the background are highlighted. It is worth noting that the task to have "identical" chromatic features is not easy to achieve in hybrid systems, where one image is acquired directly while the other is acquired as a reflection on a surface that may alter colors to some extent.

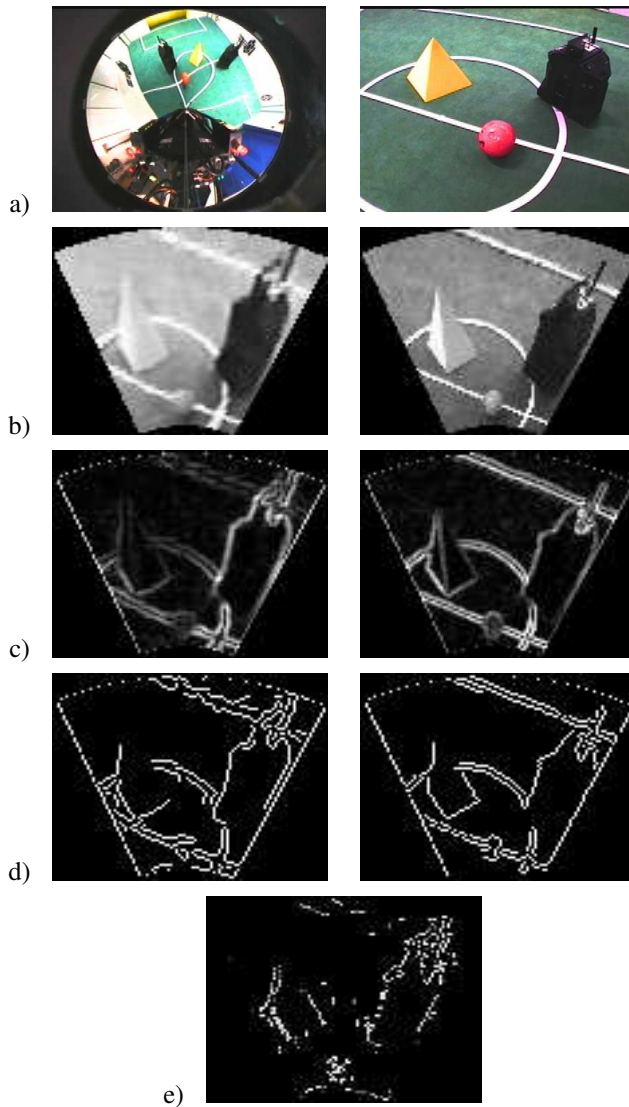


Figure 5: Obstacle detection: (a) images acquired by the hybrid vision sensor; (b) the IPT of the spatial region in (a) common to both images; (c) results of edge detection applied to (b); (d) result of the ridge extraction from (c); (e) difference between the two images in (d).

The white regions that can be observed in the difference image, that represent areas where an obstacle may be present, derive from two kinds of disparity that can be found in stereo image pairs. If they derive from a lateral displacement of the two cameras, they are located to the left and/or right of obstacle projections in the IPT transformed images; because of this, both approaches used to obtain binary difference images considered above provide very similar results. When a vertical displacement of the two cameras occur instead, such regions are located above and/or below the obstacle projections.

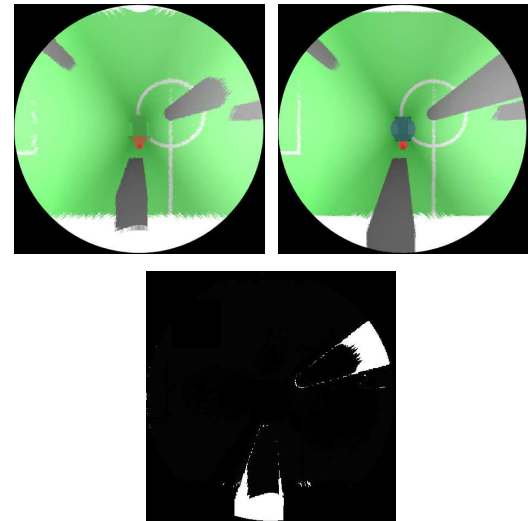


Figure 6: Above: simulated results obtained by a coaxial fully-omnidirectional system. The two IPT images (upper sensor on the left, lower on the right) of a simulated RoboCup environment. Below: the difference image that can be obtained with the coaxial configuration. The virtually null lateral disparity can be clearly noticed.

From these considerations and using other kinds of information (e.g. color) it is possible to tell regions that are certainly free from regions that may be occupied by obstacles. Figure 5 shows the results that can be obtained at the end of each step.

To give a flavor of the potential of the algorithm when applied to a fully-omnidirectional stereo device, in figure 6 the difference images is shown, which was obtained by IPT-transforming the (simulated) images taken from the two sensors, and subsequently computing and pre-processing the difference between the two images. In particular, the results of the difference between the self-reflections of the robot onto the two mirrors have been removed.

6. Discussion

In this paper we have described a Hybrid Omnidirectional/Pin-hole Sensor (HOPS) and a general framework within which the IPT is used to allow for porting the "quasi-3D" (q3D) class of stereo algorithms from traditional stereo systems to omnidirectional or partially-omnidirectional ones.

The joint use of a standard CCD camera and of an omnidirectional sensor provides HOPS with their different and complementary features: while the CCD camera can be used to acquire detailed information about a limited region of interest ("foveal vision"), the omnidirectional sensor provides wide-range, but less detailed, information about the surroundings of the system ("peripheral vision"). HOPS,

therefore, suits several kinds of applications as, for example, self-localization or obstacle detection, and makes it possible to implement peripheral/foveal active vision strategies: the wide-range sensor is used to acquire a rough representation of a large area around the system and to localize the objects or areas of interest, while the traditional camera is used to enhance the resolution with which these areas can then be analysed. The different features of the two sensors are very useful for a combined exploitation in which information gathered from both the sensors is fused, allowing extraction of "2D augmented information" from the observed scene by means of IPT.

The IPT implementation that has been proposed allows for a fully-automatic calibration of the sensor, and for a very efficient derivation and subsequent use of the mapping function, implemented through a look-up table. An algorithm for obstacle detection, based on such an implementation of the IPT has been briefly presented to show the effectiveness of the approach. One of the most noticeable features of this approach is the cancellation of "false obstacles" lying on the IPT reference plane: shadows projected on the floor, spots, drawings or bi-dimensional objects lying on the floor, which can appear in the acquired images and can be mistaken with obstacles by a monocular vision system because of their texture, color, etc. can be easily removed by IPT.

The application of the look-up table is the only overhead imposed on the algorithms by the use of IPT, with respect to their 'standard' implementation. This, along with an MMX-optimization of the code, has made it possible to achieve real-time or 'just-in-time' performance, allowing the algorithm to track objects that move with a relative speed up to over $2m/s$ on recent mid-top class PCs.

Acknowledgements

This work has been partially supported by ASI under the "Hybrid Vision System for Long Range Roving" grant and by ENEA under the "Intelligent Sensors" grant.

References

- [1] Benosman, R. and Kang, S. B., editors. *Panoramic Vision: Sensors, Theory and Applications*. Monographs in Computer Science. Springer-Verlag, New York (2001).
- [2] Gutmann, J. S., Weigel, T., and Nebel, B. Fast, accurate, and robust selflocalization in polygonal environments. *Proc. 1999 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (1999) 1412–1419.
- [3] Clérentin, A., Delahoche, L., Pégard, C., and Brassart-Gracsy, E. A localization method based on two omnidirectional perception systems cooperation. In *Proc. 2000 ICRA. Millennium Conference*, vol.2 (2000) 1219–1224.
- [4] Sogo, T., Ishiguro, H., and Trivedi, M. N-ocular stereo for real-time human tracking. In Benosman, R. and Kang, S. B., editors, *Panoramic Vision: Sensors, Theory and Applications*, Monographs in Computer Science, chapter 18. Springer-Verlag, New York (2001) 359–376
- [5] Mallot, H. A., Bühlhoff, H. H., Little, J. J., and Bohrer, S. Inverse perspective mapping simplifies optical flow computation and obstacle detection. *Biological Cybernetics*, vol.64 (1991) 177–185.
- [6] Onoguchi, K., Takeda, N., and Watanabe, M. Planar projection stereopsis method for road extraction. *IEICE Trans. Inf. & Syst.*, vol. E81-D n. 9 (1998) 1006–1018.
- [7] Adorni, G., Cagnoni, S., and Mordonini, M. An efficient perspective effect removal technique for scene interpretation. *Proc. Asian Conf. on Computer Vision* (2000) 601–605.
- [8] Adorni G., Cagnoni S., Carletti M., Mordonini M., Sgorbissa A. Designing omnidirectional vision sensors, *AI*IA Notizie*, vol. 15 n. 1 (2002) 27-30.
- [9] Adorni G., Bolognini L., Cagnoni S., Mordonini M., A non-traditional omnidirectional vision system with stereo capabilities for autonomous robots, In F. Esposito (ed.) *AI*IA 2001: Advances in Artificial Intelligence. 7th Congress of the Italian Association for AI, Bari, Italy, September 2001: Proceedings*, Springer, LNAI 2175 (2001) 344-355.