

Obtaining range information with an omnidirectional vision system

Fabiano Rogério Corrêa
Department of Mechatronics Engineering and Mechanical Systems
Escola Politécnica of the University of São Paulo
fabiano.correa@poli.usp.br

Cláudia C. G. Deccó
Department of Mechatronics Engineering and Mechanical Systems
Escola Politécnica of the University of São Paulo
ccgdecco@usp.br

Jun Okamoto Jr.
Department of Mechatronics Engineering and Mechanical Systems
Escola Politécnica of the University of São Paulo
jokamoto@usp.br

Abstract. *Autonomously guided mobile robots are dependent on sensor information for internal decision making processes. Important information for the robot is the distances between its sensors and the objects in its working environment. One way of obtaining this information is through the use of stereo vision which allows the determination of depth (range) information from two or more images. Stereo can be achieved with omnidirectional vision systems that produce panoramic images with a field of view of 360º around the camera. Basically three steps compose the process of obtaining depth information through a pair of images: features extraction, matching and triangulation. This paper presents the method of obtaining range information from a pair of omnidirectional images. It uses a CCD camera and a hyperbolic mirror mounted on a mobile robot. The environment is considered static and the robot with the camera moves to acquire the stereo pair. In this work the processing is done directly on the omnidirectional images showing effective results for its utilization for an autonomous mobile robot.*

Keywords: *omnidirectional vision, stereo vision, image matching, autonomous mobile robot, robot mapping.*

1.Introduction

It is fundamental in the design of an autonomous mobile robot to capacitate the machine to acquire information through its own sensors and to elaborate an internal representation of the world for planning actions and making decisions. At least the robot must extract information of its sensors and interpret it in a manner that it can judge the next step to take. Sensors could be of many kinds as for example vision systems.

Stereo systems play an important role in the applications for mobile robots. It is used to obtain depth information, which can be used to avoid obstacles, localize itself and navigate. In the most common case, we must determine the distances between the some objects in the world and the camera from a pair of 2-D images.

The processes involving stereo vision are basically compounded of three steps: feature extraction, matching and triangulation. A feature is extracted of the first image and one must search for the correlated feature in the second image. With this correlated feature the triangulation is used to obtain the distance of the system to this feature in the world.

The features used in stereo can be points, lines or an area of the image. Algorithms of image processing are applied to the images to extract these features.

As part of the matching process there is the search for the correspondent feature. To save computational cost in this search some constraint must be used. The constraint can be a physical or a geometrical one. Among the geometrical constraints the epipolar geometry is the mostly used.

The triangulation algorithm uses the geometric properties of the catadioptric system (i.e., a system formed by a camera and a mirror) to calculate the distances between the real object whose represented characteristics was correlated in the image pair and the camera.

Faugeras (1993) and Zhang (2002) relate a series of algorithms for stereo matching in a survey paper. They explain some representative algorithms like sum of squared differences (SSD), dynamic programming, graph cuts, layered stereo and real-time correlation-based stereo.

The great problem in stereo system is the matching of the corresponding features between the image pair. Area-based and feature-based matching methods are extensively used. Wang and Hsiao (1996) make stereo by using the Sobel operator to extract features. These features are used to train a neural network in order to obtain an adaptative matcher that can generate a disparity map. The matching algorithm developed uses some constraints like epipolar constraint, ordering, geometrical and local-support to update this initial map. They obtain a good matching and demonstrate the efficiency of the trained network.

Ku et al (2001) makes stereo with multiple images taken under a general motion. They represent the SSD in terms of the depth information. By integrating the general stereo images the ambiguities in correspondence are reduced and the accuracy of the reconstruction is improved.

For omnidirectional stereo, Ollis et al (1999) shows five different configurations to perform stereo with catadioptric systems making analysis related to the field of view, range estimation and depth calculation. Then they do a discussion about stereo matching that is done using SSD function showing the necessity to incorporate the distortion of the mirror in this process. This is accomplished by generating an unwarped image that incorporates the mirror geometry.

Using a pair of images acquired with a hyperbolic mirror, Wei (2000) establishes a dense stereo correspondence by using SSD function and dynamic programming with uniqueness and ordering constraints imposed. Because the distortion of the images, he does the matching recovering the epipolar geometry. He found out that it is very important to collect enough and representative samples along the epipolar curves to achieve a good matching. It was studied and experimented the cases of sampling rate of points along the epipolar curves, the size of the strip as the support region for correlation and the separation between adjacent neighboring epipolar curves. It is showed the cases that result in a better match.

In this paper we use as features the points of the borders of objects and the epipolar constraint to reduce the search of correspondent points. The matching utilizes an SSD algorithm. The stereo image pair is obtained with a displacement of the single camera and the hypothesis of a static environment. All those processes are done directly on the omnidirectional image.

In the Section 2 we present the omnidirectional vision system and briefly discuss the manners that it can be designed. In Section 3 we explain omnidirectional stereo vision describing each step to obtain range information. In Section 4 and 5 respectively we shows the results and conclusions.

2.The Omnidirectional Vision System

An omnidirectional image is the projection of a 360 degrees field of view in a single image. This image can be produced in several ways. One way is through the composition of several images from a camera that rotates in one of its axis and captures scenes of all parts of the room. In another way several fixed cameras each pointed in one direction substitute the rotation of the camera. Also the utilization of a special lens in the camera, such as the fish-eye lens, permits to acquire a wide-angle view image in real-time. Finally, an omnidirectional image can be obtained with a simple camera and a mirror. The camera is pointed vertically toward a mirror with the optical axis of camera lens aligned with the mirror's axis. Among the many possibilities for mirror profiles that can be used, the conical or convex such as spherical, parabolic and hyperbolic are the simplest. A 360 degrees image is obtained with this compact system known as catadioptric system.

Although there exists many configurations to obtain an omnidirectional image it is convenient that the system has a single center of projection, or a single viewpoint. This enables the generation of any desired image (perspective or panoramic) projected on any designated image plane. For a rotation camera a single center of projection can be obtained by an exact alignment of the camera axis and the center of the projection of the camera. For the case of multiple cameras this is achieved when all viewpoints are coincident. For the catadioptric system this can be satisfied when all rays that intercepts the mirror surface are reflected passing through the focus of the mirror. This is achieved with the hyperbolic and parabolic mirrors, but with the latter it is necessary the use of an orthographic projection camera (Svoboda et al, 1998).

A camera and a hyperbolic mirror compose the omnidirectional vision system developed in this work. The camera and the mirror are fixed in an acrylic cylinder mounted on the robot as shown in the Fig. (1). The hyperbolic mirror is milled in aluminum in an ultra-precision lathe machine. The acrylic doesn't cause any major distortion in the resolution of the images. This system is mounted on a mobile robot to be used for navigation. The system was designed to produce a self-alignment of the mirror with axis of the camera.

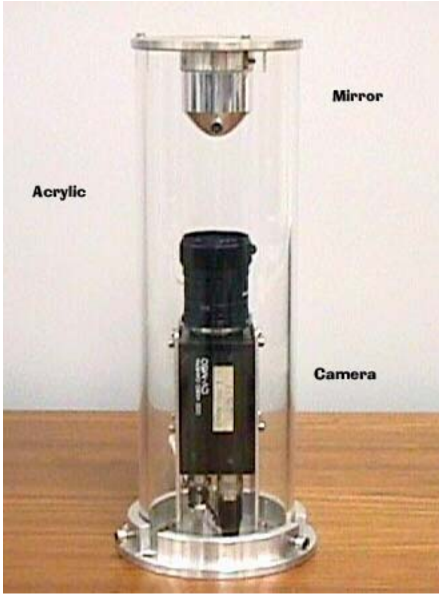

Figure 1. The omnidirectional vision system mounted on the robot.

3.Omnidirectional Stereo Vision

In the binocular stereo vision a pair of cameras acquire two images of the same scene to obtain depth information. First, some characteristics of one image are selected for the search in the second one. Then, geometrical constraints between corresponding features are used to match them. Finally, triangulation process is used to obtain the 3-D information of an object (Faugeras, 1993).

The process of feature extraction consists on finding specific characteristics, which will be used for the matching process. These features can be an area or points in the image with specific properties such as corners or borders of an object. Given a feature in the first image it is necessary to find the correspondent feature in the second image. This process is called matching. The positions in the images of the features matched and the displacement of the cameras are used to find the positions of the points in space. This is the triangulation process.

In the motion stereo technique a single moving camera acquires the images used for the system. Whether the motion is parallel to the optical axis of the camera (axial stereo) or perpendicular to it (lateral stereo) the matching process obtain good matching results (Ku et al, 2001). However, it is quite difficult to maintain the camera motion parallel or perpendicular to the optical axis in many real situations. There are also many stereo techniques using multiple images.

In this paper we presents the stereo vision system that acquire two images taken with a small displacement of a single camera. Some characteristics are extracted and the sum of squared differences (SSD) is used to find the corresponding points of them.

3.1.Feature Extraction

Among the features that can be used for the purpose of calculate depth information from a pair of images are the points, lines and areas. Utilizing the appropriated image processing algorithms as gradient or others masks one can extract the intended features.

The features used for the stereo algorithm in this work are the edges of the objects. These points are extracted with an algorithm of image processing known as SUSAN (Smith, 1997). The SUSAN (Smallest Univalue Segment Area Nucleus) algorithm is a low level processing, where a circular mask is passed over the image to find the amount of points that have the same gray level as the center of the mask. With this value it is possible to determine if the point is in the edge of an object. When the circular mask is centered in a pixel it determines the numbers of neighborhood pixels with the same intensity of light (based on a threshold that compensates for the image contrast) of its center is done. If half of the pixels in the mask are of the same intensity of the center of the mask signify that this pixel is on the border of an object (Fig. (2)).

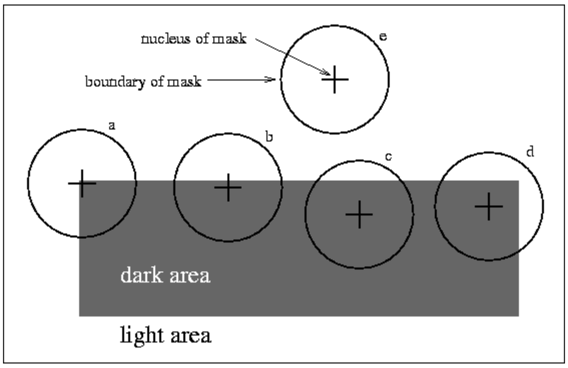

Figure 2. The principle of edge extraction of the SUSAN algorithm (Smith, 1997).

Figure (3) shows an example of the results using the SUSAN algorithm in an omnidirectional image. The red dots are the extracted corner points. The algorithm needn't any change to work on this kind of image because it just detects variation in the gray level of the pixels independently of the form of the image (warped or unwarped).

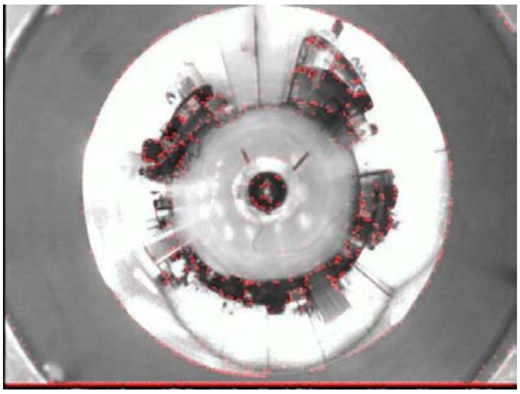

Figure 3. Corner points extracted in the first image.

3.2.Matching

The matching between the features extracted in the first and second images of the stereo pair is the most sensible part of the system. A constraint is necessary to reduce and simplify the search for the matching features in the other image of the pair thus reducing the computational cost. One of the most used constraints is the epipolar geometry. The epipolar constraint determines that the matching feature is located in a determined geometric locus (a curve), thus decreasing the search dimension from two to one.

In the omnidirectional case, there exists an epipolar plane (plane π) formed by three points: the focus (points F_1 and F_2) of the mirror in the two positions where the images are acquired, and the position of the feature (point P) in the real world. This plane intercepts the mirrors and the projection of the intersection defines in the image the epipolar curve as shown in Fig. (4). Figure (5) shows the epipolar curves determined for all features extracted in the first image shown in Fig. (3).

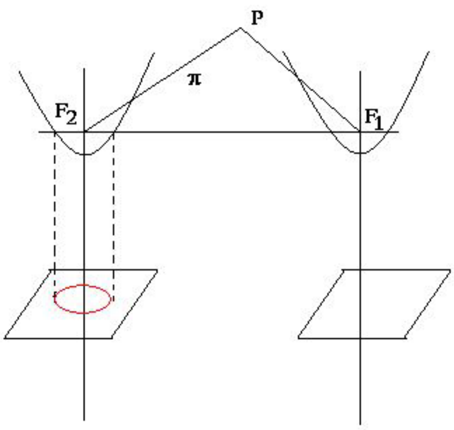

Figure 4. Omnidirectional Epipolar Geometry.

Most of the researchers that work with omnidirectional images project the image on a cylinder transforming the circular image into a rectangular one. In this new rectangular image, the epipolar curve turns to be a line. The effort of this work is to make the matching directly with the omnidirectional (circular) images.

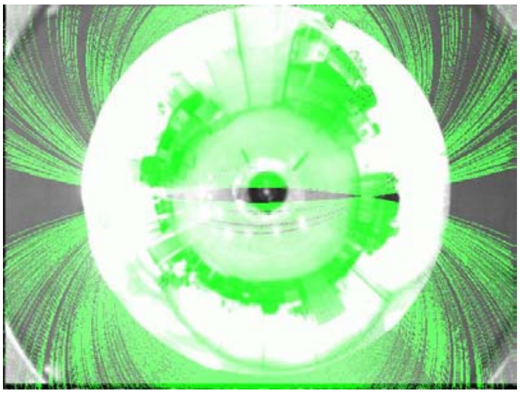

Figure 5. The epipolar curves in the second image.

There is a trade off between the distance of the cameras (baseline) and the quality of the results of the matching process. Smaller displacement between the camera positions does not change too much the illumination of the scene but reflects in a poor triangulation. Greater displacement results in good triangulation but also in large changes in illumination.

For each feature extracted in the first image, the corresponding epipolar curve is calculated in the second image. A rectangular window-mask is then passed with its center positioned on the epipolar curve of the second image to find the correspondent point for the feature extracted in the first image. The size of this window is a critical choice. It is another trade-off between the size of the window and the information necessary for matching. A small-window in a smooth region of the image cannot provide sufficient information to find the exact correspondent pixel whereas a large window could generate ambiguities problem or false matching.

The search is performed using the SSD (Sum of Squared Differences) between the pixels inside the mask in the two images, according to the following expression:

$$E(i_0, j_0) = \sum_{(i,j)\in W_k(i_0,j_0)} \left(I_1(i,j) - I_2(i,j)\right)^2 , \tag{1}$$

where E is the result of the SSD; I_1 and I_2 are the intensity level of the pixels of image 1 and 2 respectively, i_o and j_o are the indexes of the central pixel of the mask in the second image, and i and j are the indexes of the pixels inside the masks on images 1 and 2. Note that the mask in the first image is centered on the feature extracted and the mask in the second image is centered in all pixels along the epilolar curve. Obviously for each pixel in the epipolar curve on the second image it is calculated a value for E. The minimum value along the epipolar curve indicates the correct match, i.e., the position in the second image that matches the feature of the first image. The idea behind the SSD is to increase through the squared difference the differences between the scenes inside the masks.

Figure (6) shows the matching points of the second image that correspond to the features extract in the first image. Compare Fig. (3) and (6) to see the pair of features matched in both images.

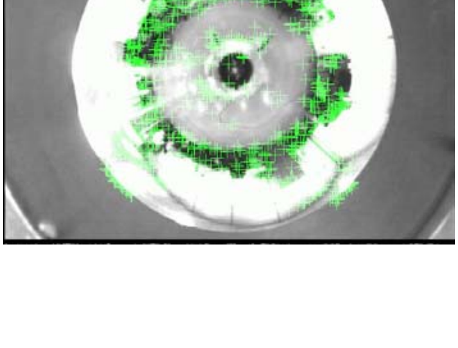

Figure 6. The matching points.

3.3. Triangulation

Having the feature extracted in both images it is possible to find the distance between that feature in the real world and the camera.

In the triangulation one must make the inverse process of the light that activated the photo sensors in the camera's CCD. With the pixels in the image we could calculate the direction of the distance vector starting from the focus of the mirror and pointing to the object in the real world.

Firstly, from the images and applying the geometric properties of the catadioptric system, we found the vector with the direction of the feature in the space starting at the center of the mirror. Secondly, with the two vectors the distance is calculated utilizing triangle similitude.

To find the vectors pointing to the real object in the space through the point in the image, the problem to be solved is the system that models the intersection between a line (the ray of light) and a hyperbole (the mirror surface).

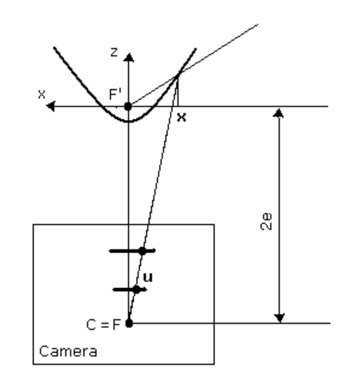

Figure 7 – Catadioptric system

The line **s** that represents the ray of light that reflects in the mirror surface and goes to the focus of the camera (modeled as a pin-hole camera) thus marking a pixel in the image plane is,

$$s : z = \frac{1}{u}x - 2e \tag{2}$$

where **u** is the squared root of the sum of squared normalized coordinates of the pixel in the image plane and **e** is a parameter that depend of the mirror geometry (Fig. (7)). The equation was deducted for a plane y = 0 although the result is valid to the whole system with a substitution of variable.

The hyperbole **h** is the surface mirror and is represented by,

$$h : z = \sqrt{a^2\left(1 + \frac{x^2}{b^2}\right)} - e \tag{3}$$

where **a** and **b** are parameters of the mirror's geometry (see Svoboda98 for a detailed model) and $e = \sqrt{a^2 + b^2}$.

The intersection of this line with this hyperbole give us the vector that points to the real point in the world, and is obtained through a substitution of **z** of the equation (2) in (3). Solving for **x** we obtain:

$$x = \frac{\frac{2e}{u} \pm \sqrt{\frac{4e^2}{u^2} - 4\left(\frac{1}{u^2} - \left(\frac{a}{b}\right)^2\right)\left(e^2 - a^2\right)}}{2\left(\frac{1}{u^2} - \left(\frac{a}{b}\right)^2\right)} \tag{4}$$

where x is the projection on plane parallel to x-y of a vector. This vector begins in the mirror's focus and ends in the surface of the mirror.

With the vectors and its projections in the coordinate system, it is possible to draw a triangle and thus apply the formulas of similitude between triangles to determine the distance of point P projected in the plane x-y to the line of displacement that connect F1 and F2.

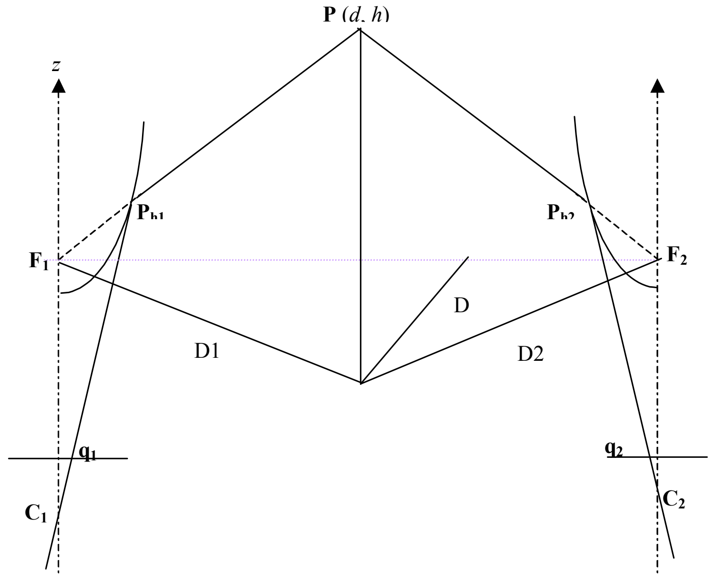

Figure 8. Similitude of triangles.

The similitude of triangles is obtained through comparing the first triangle (with sides D1, D and the line starting in F1 and ending in D) and the second triangle (with sides D2, D and the line starting in F2 and ending in D) with the projections of the vectors encountered previously (x e y). Thus the equations below are calculated:

$$D_i = \sqrt{D^2\left(1+\frac{x_i^{\ 2}}{y_i^{\ 2}}\right)} \tag{5}$$

$$D = \frac{y_1 y_2}{x_1 y_2 + x_2 y_1} \tag{6}$$

4.Results

This stereo vision system was tested with a virtual environment created in POV-Ray software. In a virtual environment, it is possible to control the illumination and others noises that cause a misinterpretation of the results of the algorithms of the system. Results in a real environment were also included in the tests and are reproduced below.

The images of the stereo pair were taken with a displacement of 200 mm of the first to the second position of the system. Smaller displacements make the triangulation mathematically and computationally difficult because of the small angle in the triangle utilized to determine the distances and in bigger displacements the distortion of the object because of the mirror is so great that the matching is not accurate.

Below were showed images of the processing of the stereo system. Figure (9) shows points in the edges of a cube. This points were chose from a list of points extracted by the SUSAN algorithm. We use in this case a 12 x 12 window in the SSD algorithm.

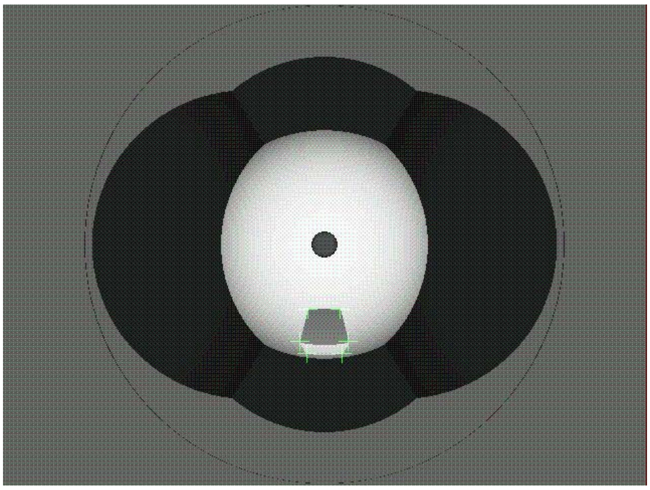

Figure 9. Points extracted in the first image.

Figure (10) shows the epipolar constraint that is the locus where to search for the matching points in the second image of the pair. Note that it passes in the edges of the cube to be found.

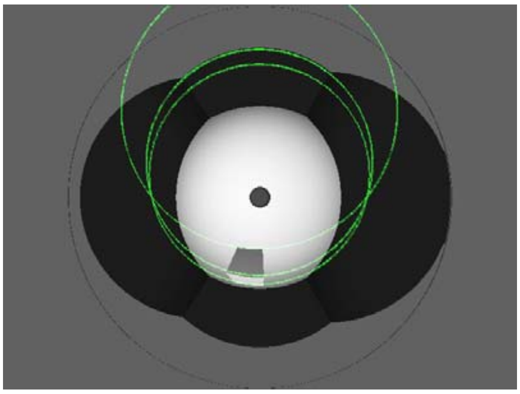

Figure 11. Epipolar curves showing the candidates for matching.

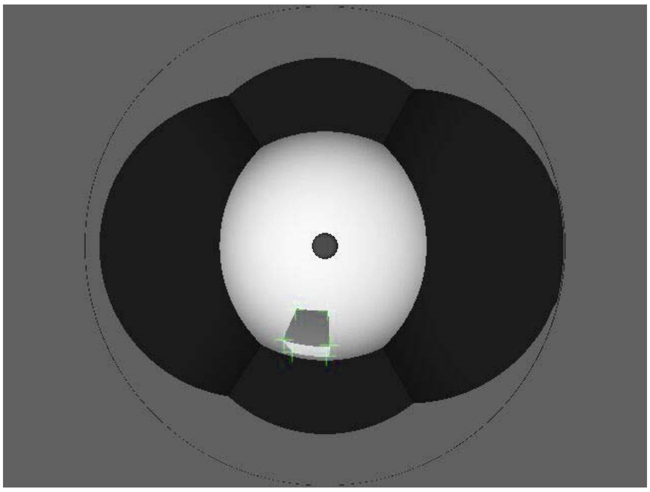

Figure 10. The matching.

Figure (11) shows the matching of those points extracted in the first image and Fig. (12) is a map (superior view) based on the triangulation calculus. Note that the edges of the frontal face should be projected in the same point in each side of the cube. Yet note that in the omnidirectional images, acquired through a camera and a mirror, the objects appear in the image in opposition of the real side, as in an ordinary mirror. The black dots represent the vertices of the cube; this figure is an illustrative way to see the detected points.

Figure 12. The distance map.

The test was realized in a real environment Fig. (13) is a scene of our lab. The calculus by the stereo algorithm was focused in a box (circled in red in the Fig. (13)), which is a 1-meter of the robot (in center of the image). The distance obtained by the system based in points detect in one face of the box is 808.6 mm in average.

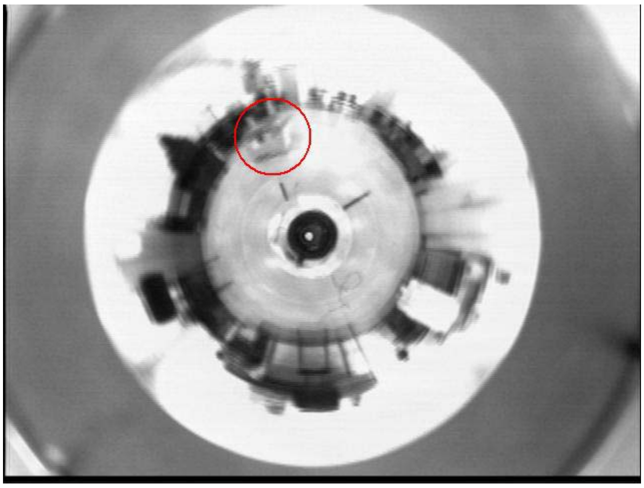

Figure 13. A real environment.

5.Conclusion

The omnidirectional stereo vision sensor can gather information of the robot working environment. To calculate distances from objects to the camera an algorithm with the following three steps is used: feature extraction, matching and triangulation. The SUSAN algorithm implemented for feature extraction was able to find an amount of points in the image sufficient to construct a dense depth map. The matching was performed directly in the omnidirectional image by using the epipolar constraint and the SSD function. The epipolar constraint was the choice to restrain the space of search and could be utilized directly in omnidirectional images. The matching process is sensible to the size of the window but is capable of finding good matches. The triangulation works better when the displacement between the two images is not too large avoiding thus large differences of the projection of objects in both images. The error in the distances calculated is acceptable for applications like robot navigation.

The results show that working directly with the omnidirectional image it is possible to obtain reliable data and to save computational cost. Some improvements could be made in the system including applying in the matching search other constraints than the epipolar geometry.

The problems encountered are regarded primarily to the matching algorithm that must be adjusted but this results are adequate to tasks of autonomous mobile robots as navigation relying in the vision system to obtain range information to detect and avoid obstacles.

6. References

Faugeras, O., 1993, “Three-dimensional Computer Vision: A Geometric Viewpoint”, MIT Press.

Faugeras, O., Hotz, B., Mathieu, H., Viéville, T., Zhang, Z., Fua, P., Théron, E., Moll, L., Berry, G., Vuillemin, J., Bertin, P. and Proy, C., 1993, "Real time correlation-based stereo: algorithm, implementations and applications", Institut National de Recherche en Informatique et en Automatique, INRIA, 49 p.

Gonzalez, R. C., Wintz, P., 1993, "Digital image processing", Addison-Wesley, Reading, MA.

Ku, J. S., Lee, K. M. and Lee, S. U., 2001, "Multi-image Matching for a General Motion Stereo Camera Model", Pattern Recognition, Vol. 34, pp.1701-1712.

Naccache, F. F. and Hachiya, R. K., 2001, "Visão estéreo a partir de seqüências de imagens geradas por uma camera panorâmica central com espelho hiperbólico". Trabalho de formatura, Escola Politécnica da Universidade de São Paulo.

Ollis, M., Herman, H. and Singh, S., 1999, "Analysis and Design of Panoramic Stereo Vision Using Equi-Angular Pixel Cameras", CMU-RI-TR-99-04, The Robotics Institute, Carnegie Mellon University. 5000 Forbes Avenue, Pittsburgh, PA, 15213, USA, 43 p.

Okutomi, M. and Kanade, T., 1993, "A multiple-baseline stereo". IEEE Transactions on pattern analysis and machine intelligence, vol. 15, no. 4.

Smith, S. M., Brandy, J. M., 1997, "SUSAN - A new approach to low level image processing". *Int. Journal of Computer Vision*, 23(1):45—78.

Svoboda, T., Pajdla, T., Hlavác, V., 1997, "Central Panoramic Cameras: Geometry and Design", Technical Report Nº K335/97/147, Czech Technical University, Department of Control Engineering, December.

Svoboda, T., Pajdla, T., Hlavác, V., 1998, "Epipolar Geometry for Panoramic Cameras". In fifth European Conference on Computer Vision, Freiburg, Germany, pp. 218-232.

Wang, J. H. and Hsiao, C. P., 1996, "Stereo Matching by Neural Network That Uses Sobel Feature Data", IEEE International Conference on Neural Networks, Vol. 3, pp.1801-1806.

Wei, S. K., 2000, "Stereo Matching of Catadioptric Panoramic Images", Center for Machine Perception, Dept. of Cybernetics, Faculty of Electrical Engineering, Czech Technical University, CTU-CMP-2000-08.

Zhang, C., "A Survey on Stereo Vision for Mobile Robots" Dept. of Electrical and Computer Engineering, Carnegie Mellon University. 5000 Forbes Avenue, Pittsburgh, PA, 15213, USA, czhang@andrew.cmu.edu.