An Omnidirectional Vision System for Outdoor Mobile Robots

Wen Lik Dennis Lui[1] and Ray A. Jarvis[2]

Intelligent Robotics Research Centre, Monash University,
Clayton Victoria 3168, Australia,
dwllui@gmail.com[1], Ray.Jarvis@eng.monash.edu.au[2]

Abstract. The advancements made in microprocessor and image sensory technology have made inexpensive, fast and robust computers and high resolution cameras widely available. This opens up many new possibilities as researchers can now take advantage of the rich visual information of the environment provided by the vision system. However, conventional cameras have a limited field of view, which is a constraint for certain applications in computational vision. As a result, the use of omnidirectional vision systems have become more prevalent within the robotics community in recent years. In this paper, a novel variable multi-baseline stereo omnidirectional vision system and its algorithms intended for outdoor navigation will be presented.

Key words: Omnidirectional Vision, Catadioptrics, Stereovision

1 Introduction

Vision has always been regarded as one of the most important form of information not only for robots, but for humans as well. The human vision system is so powerful that it is able to perceive depth, segment out and classify objects in the scene and perform other high level perceptions, all in real time. As such, it has been long since researchers are inspired to mimic the human vision system by using computers and cameras. Similar to human eyes, conventional cameras have a limited field of view. However, the capability to process such information for higher levels of perception on a modern computer is still far less efficient and accurate as compared to that of the human brain. As such, omnidirectional vision systems have been gaining popularity in the robotics community as it sidesteps the commonly encountered *windowing* problem by capturing a much wider and continuous field of view. Besides, in the context of autonomous robot navigation, omnidirectional vision systems work well for both appearance-based and landmark-based localization approaches since most features in the scene will still be visible after some small arbitrary amount of motion.

In literature, there is an exhaustive set of techniques that can be used to create an omnidirectional vision system. A popular method is to combine a camera with mirror(s). This class of omnidirectional vision system is more commonly known as a catadioptric camera system. Generally, these mirror(s) is/are either

planar or curved. The proposed omnidirectional vision system in this paper are actually catadioptric systems which utilize curved mirrors. As such, the remainder of this introduction will be dedicated to this class of omnidirectional vision system. Readers interested to learn more about other classes of omnidirectional vision systems should refer to the book compiled by Benosman, R. and Kang, S. B. [1] for more details.

Catadioptric systems can be further classified into central and non-central systems. Central catadioptric systems possess a single effective viewpoint whereas non-central catadioptric systems have multiple effective viewpoints. The central catadioptric systems have been thoroughly studied by Baker, S. and Nayar, S. K. [2] and Svoboda, T. and Pajdla, T. [3]. In [2], a complete class of single-lens single-mirror catadioptric sensors that have a single effective viewpoint has been derived. In addition, a single effective viewpoint was described to be desirable since it permits the generation of geometrically correct perspective images from the original image captured by the system. This is then followed by the derivation of the epipolar geometry for central catadioptric cameras in [3]. The epipolar geometry is a rather valuable property for mobile robotic applications as it eases the search for corresponding points in two images taken by the central catadioptric system. Subsequently, the surrounding environment can be represented in 3D by means of triangulation. Having a 3D perception of the outdoor environment is particularly useful as it is often a prerequisite for navigation and path planning purposes.

Unfortunately, most of the catadioptric systems are non-central. This is because only a limited number of camera-mirror combinations can be classified as central catadioptrics. In addition, a slight misalignment during system setup and the vibration experienced during robot motion might transform it into a non-central system. For the case of non-central catadioptric systems, Swaminathan et al. [4] have described the locus of the multiple effective viewpoints as caustics. An in-depth analysis of caustics for catadioptric cameras with conic reflectors is provided and a calibration technique for conic catadioptric cameras with known motion is proposed. As for the purpose of extracting 3D information from two images taken from the same scene at different positions, Gonçalves, N. and Araújo, H. [5] proposed an estimation technique based on the geometrical relationships of the mapping of 3D scene points to image points on quadric mirror surfaces. On the other hand, Mičušík, B. and Pajdla, T. [6] proposed a suitable non-central model approximated from a central model in order to obtain a projection model for a spherical mirror. Consequently, this directly implies that a completely different model might be required for a different curved mirror since no single model can be general enough to cater for all types of mirrors.

The main contribution of this paper is the development of an omnidirectional vision system (or more specifically in this case, a non-central catadioptric vision system), which utilizes equiangular mirrors, intended for outdoor robot navigation. The proposed system exploits the properties of the equiangular mirror in order to estimate 3D positions of scene points in the captured image via stereovision techniques. A brief system overview can be found in Sect. 2 which is then

followed by the description of the camera calibration technique in Sect. 3. Then, the underlying stereovision techniques implemented will be presented in Sect. 4 and the results will be presented and discussed in Sect. 5. Finally, this paper will be concluded with possible future extensions in Sect. 6.

2   System Overview

The ActivMedia Pioneer P3-AT is used as the main research platform in this project. It is a highly versatile all-terrain robotic platform used in many laboratories across the world. The two catadioptric camera systems, each consisting of a Canon Powershot S3 IS digital camera pointing upwards to an equiangular mirror, are aligned vertically as shown in Fig. 1. The Canon Powershot S3 IS can capture pictures at resolutions up to 6 megapixels. A high resolution image will be useful for outdoor environments since distant features will appear much smaller in the image. However, there is a fair tradeoff between computational workload and accuracy, and striking this balance is dependant on the system developer.
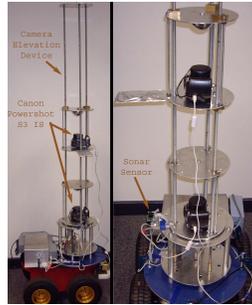
Fig. 1. (Right)The Pioneer P3-AT equipped with a camera elevation device (with two non-central catadioptric systems, each consisting an equiangular mirror and a high resolution digital camera) which allows the robot to capture images at different baselines and (Left) a close up view of the location of the sonar sensor which is used to measure the effective baseline

This robotic platform is equipped with a novel camera elevation device as shown in Fig. 1. Due to this unique device, this robot is called the *Eye-Full Tower*. This device is used to translate the upper catadioptric system vertically such that the effective baseline can be varied. It works by means of controlling a DC motor via a parallel port. The DC motor, which is connected to a gearing system (enclosed at the bottom of the device), will then rotate the three long and threaded rods which is connected to the brass bits attached on top of the upper catadioptric system (clearly shown in Fig. 1 (Right)). The camera elevation device can translate the upper catadioptric system from a range of 30cm to 100cm measured from the lower catadioptric system. The effective baseline will

be constantly monitored by a sonar sensor which is accurate up to 0.5cm. Since this project has chosen the equiangular mirror over other curved mirrors, the following subsection will describe the properties and highlight the advantages of utilizing this mirror.

2.1   Equiangular Mirror

Equiangular mirrors are a special class of curved mirrors which possess a linear relationship between the ratio of change in radial angle with the angle of elevation as shown in Fig. 2. The mirror used in this project is designed by Chahl, J. S. and Srinivasan, M. V. [7]. One of the main advantages of using this mirror is the relatively simple algorithm (Cartesian to Polar coordinate transform) which can be used to unwarp the original image to a more intuitively understandable form. In addition, this unwarpping process is rather useful as they sidesteps the scaling problem of image neighborhoods during the establishment of stereo correspondences via an area-based method.
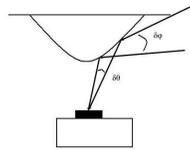
Fig. 2. Elevation angular magnification α (constant), the ratio of change in radial angle $\delta\theta$ with respect to change in angle of elevation $\delta\varphi$

It is also important to observe that pixels lying on the same line propagating outwards from the centre of the mirror towards the rim of the mirror in one image can be found on the same line in another image taken at a different vertical position. This property of the mirror, more commonly known as the epipolar line, was first reported by Ollis, M. et al [8] and various system configurations have been analyzed. In this project, a vertical configuration is adopted in order to exploit this property of the mirror. Assuming that the lower catadioptric image is always taken as the reference image, the corresponding feature in the upper catadioptric image will always be closer to the centre of the mirror in the image as illustrated in Fig. 3 due to the change in radial angle and angle of elevation of the same scene point.
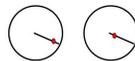
Fig. 3. (Left) Feature (in red) in lower catadioptric image and (Right) corresponding pixel on the upper catadioptric image

Before the underlying stereovision techniques are presented, a new camera calibration technique which is suitable for this system will be described in the following section.

3   Camera Calibration

Standard camera calibration techniques involve the computation of intrinsic and extrinsic parameters of the camera system. However, since standard catadioptric camera calibration techniques are dependant on the type of mirror employed, a new calibration technique is proposed for this system. The idea of the proposed technique is based on the geometrical properties of the catadioptric system. In contrast to standard calibration techniques which compute the intrinsic and extrinsic parameters of the camera system, the proposed method attempts to compute the respective angle of elevations of the mirror with respect to the radial distances from the centre of the mirror in the image.

The calibration grid is wrapped around the catadioptric system as illustrated in Fig. 4 (Top Right and Bottom Left). Subsequently, by making use of the available geometrical information of the catadioptric system labeled in Fig. 4, the following equations can be used to compute the angle of elevation for the pixel of interest in the image (in this case, the angle of elevation at the location of the red star) which happens to fall on the epipolar line propagating outwards from the centre of the mirror towards the mirror rim in the image at angle, θ_e, which ranges from 0° to 360°. (Please take note that as reference, the epipolar line at $0^o/360^o$ falls on the same line which propagates vertically from the centre of the mirror to the mirror rim in the image).
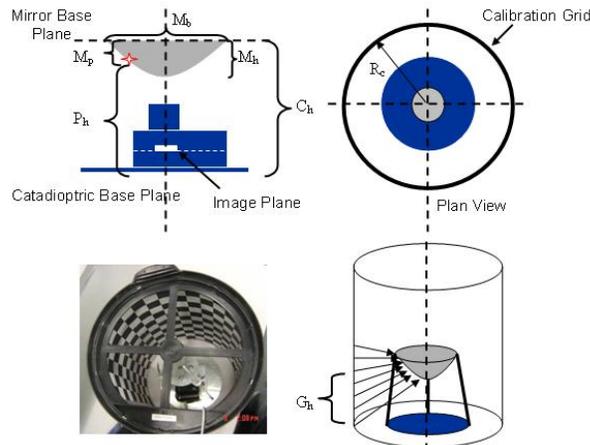
Fig. 4. (Top Left) Cross section of catadioptric structure, (Top Right) Plan view of calibration grid wrapping around catadioptric structure, (Bottom Left) Actual calibration configuration and (Bottom Right) 3D model of the actual calibration configuration

The following equation computes the parameter, P_h, which measures the height of the point of interest on the mirror with respect to the catadioptric base,

$$P_h = C_h - (M_h - M_p) \tag{1}$$

where M_p is the height of the point measured from the mirror base which can be computed by solving the following equation which describes the profile of the curved mirror when the elevation angular magnification factor is $\alpha = 7$,

$$\left(x^2 + y^2\right)^2 - 8x^2y^2 = r_o^4 \tag{2}$$

where (x,y) desribes the profile of the mirror in Cartesian coordinates, with y chosen to represent the horizontal direction of the mirror profile while x represents the vertical direction, and $r_o = 79.4$ for this mirror. The general quartic equation can then be described as,

$$a_o x^4 + a_1 x^3 + a_2 x^2 + a_3 x + a_4 = 0 \tag{3}$$

If $a_3 = a_1 = 0$, the general quartic equation will become a biquadratic equation with the general form as follows,

$$a_o x^4 + a_2 x^2 + a_4 = 0 \tag{4}$$

By simpifying equation (2) into the form of (4) and by letting $z = x^2$ in (2), the roots can be obtained by solving the simple quadratic equation,

$$z^2 - 6y^2 z + y^4 = r_o^4 \tag{5}$$

where y, representing the horizontal direction of the mirror profile, can also be defined as the actual radial distance (in mm) where the incident light ray reflects off the mirror which is measured from the centre of the mirror and can be computed as

$$y = \frac{\sqrt{\left(P_x - M_{cx}\right)^2 + \left(P_y - M_{cy}\right)^2}}{\sqrt{\left(M_{rx} - M_{cx}\right)^2 + \left(M_{ry} - M_{cy}\right)^2}} \times \frac{M_b}{2} \tag{6}$$

where (M_{cx}, M_{cy}) is the pixel location of the centre of the mirror in the image, (M_{rx}, M_{ry}) is the pixel location of the mirror rim in the image and (P_x, P_y) is the location of the point of interest in the image. The roots, z_1 and z_2, of equation (5) can then be obtained by substituting appropriate values of y and r_o into the equation. Since $z = x^2$, there will be four possible solutions for x and the following solution defines the parameter M_p,

$$M_p = +\sqrt{z_1} \tag{7}$$

where

$$z_1 = \frac{6y^2 + \sqrt{(6y^2)^2 - 4(y^4 - r_o^4)}}{2} \tag{8}$$

and

$$z_2 = \frac{6y^2 - \sqrt{(6y^2)^2 - 4(y^4 - r_o^4)}}{2} \tag{9}$$

This is a manual calibration technique and it requires the manual selection of points on the calibration grid by the user as shown in Fig. 5. Points are selected based on the location of the vertices of the square grids in the image and are automatically refined by analyzing a small window surrounding the selected point. Since the square grids are uniform in size, it is then relatively easy to determine the parameter, G_h, which measures the actual height of the selected point on the calibration grid in the image from ground (assuming that both the calibration grid catadioptric camera structure are placed on the ground). With all this information available, the angle of elevation, φ, of a particular point in the image can be determined by,

$$\varphi = tan^{-1}\frac{R_c - y}{|P_h - G_h|} \quad \text{if} \quad P_h - G_h >= 0 \tag{10}$$

$$\varphi = tan^{-1}\frac{|P_h - G_h|}{R_c - y} + 90 \quad \text{if} \quad P_h - G_h < 0 \tag{11}$$

Three assumptions are made for the calibration process; (a) The base of the mirror is parallel to the image plane and catadioptric base plane, (b) the centre of the curved mirror is aligned to the centre of the catadioptric base and (c) the calibration grid is assumed to wrap perfectly around the catadioptric structure as illustrated in Fig. 4. If these three assumptions are fully satisfied, it is possible to implement a simpler algorithm for this process as the angle of elevations computed for selected points on a particular epipolar line at a particular angle can be applied to all epipolar lines at angle θ_e. Unfortunately, this is not the case. In order to improve the accuracy of estimated 3D points in the scene, angle of elevations for epipolar lines at regular intervals (defined by the intervals of the square grids of the calibration grid in the image) are computed as shown in Fig. 5. The angle of elevations (degrees) with respect to radial distance (pixels) from the centre of the mirror in the image along epipolar lines at different angles will be stored in a lookup table. To compute the angle of elevation of a point located on an epipolar line at angle θ_{e1} and radial distance R_{d1}, linear interpolation/extrapolation of the information stored in the lookup table will be performed. Linear interpolation/extrapolation is deemed suitable as angle of elevations increases proportionally with the increase in radial distance of the point in the image.

Once stereo correspondences are established, the angle of elevations can be computed based on the location of the stereo pair in the image and the 3D
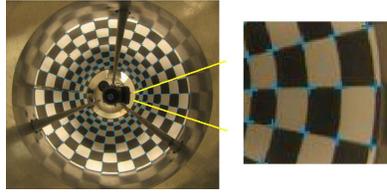
Fig. 5. Manual selection of points (in blue) on the calibration grid in the image

position of the scene point can be estimated through triangulation. Thus, the following section will describe the underlying techniques used to establish stereo correspondences.

4   Catadioptric Stereovision

Stereovision has been an active research field for many years. This phenomenon, which describes the ability of humans to perceive depth through binocular vision, was first discovered and described by Wheatstone, C. [9]. Stereovision techniques can be broadly categorized into feature-based and area-based approaches. In feature-based approaches, unique features in the image such as distinctive edges or corners [10], SIFT features [11], etc that are detected in the reference image are matched with pixels on the corresponding image. On the other hand, area-based or correlation-based approaches match pixels based on patches of pixel intensities. Area-based approaches are generally slower. However, it can produce a denser depth map (with respect to feature based approaches), which is highly desirable for robot navigation and path planning purposes.

Area-based stereovision algorithms can be further classified into local and global approaches. As defined in [12], local algorithms depend only on the intensity values within a finite window by making implicit smoothness assumptions wheareas global algorithms solve an optimization problem by making explicit smoothness assumptions in the computation of disparity. A local, area-based stereovision algorithm is implemented in this project since local approaches are much more efficient. Of course, efficiency normally trades off with accuracy and these approaches yield less accurate disparity maps.

The most popular local, area-based approaches are based on either the computation of the Sum of Absolute Difference (SAD), Sum of Squared Difference (SSD), Normalized Cross Correlation (NCC), Rank or Census. SAD is the most preferable approach and is commonly implemented on robotics applications as it is less computationally expensive. To produce a high quality disparity map, the epipolar constraint, ordering constraint and continuity constraint are applied. In addition, the popular left-right consistency constraint [13] is replaced in this system by the more efficient Single Matching Phase (SMP) method proposed by Stefano, L. D. et al [14].

In contrast to catadioptric camera systems, stereo correspondences are established after the captured image is rectified on conventional camera systems.

However, on catadioptric camera systems, establishing stereo correspondences directly on the original image will require complicated image neighborhoods for area-based methods. As described in Sect. 2.1, this problem can be avoided by unwarping the original image into a form, a rectangular image, which is more intuitively comprehensible. The epipolar lines on the original image will then transform into columns of the rectangular image and correspondences are searched for along the same column of the corresponding image.

This algorithm utilizes color images and the average of the SAD values across the three R, G and B channels are computed. A maximum disparity range is specified to limit the search range and the algorithm keeps track of the location where the minimum SAD value is found. Various thresholds such as the minimum SAD score threshold computed from the reference image neighborhood and the maximum closest SAD score threshold are applied in order to reduce ambiguity in the stereo matches. In addition, subpixel accuracy is achieved by means of parabolic interpolation. By specifying three points (the location and value of the neighbors of the minimum SAD score and the SAD score itself), a parabolic curve can be fitted onto the points and the local minimum of the curve is the actual minimum SAD score.

Finally, the algorithm will utilize the locations of these stereo pairs (converting this back into the coordinates of the original image), the effective baseline and information from the lookup table to perform triangulation. The resulting 3D position of the scene point will be described in the form of (x_p, y_p, θ_e) where θ_e describes the direction of the epipolar line in the original image and x_p and y_p are coordinates on the plane defined by the actual location of the stereo pair on the mirror. (x_p, y_p, θ_e) can then be converted to Cartesian coordinates (x,y,z) by simple trigonometric relationships. The following section will illustrate results produced by this system.

5   Results and Discussion

The current system utilizes color images of 640 x 480 in resolution. Although the Canon Powershot S3 IS can capture images up to 6 megapixels in resolution, however, processing such large images will severely slow down the algorithm. Similar to the work in [14], this algorithm reduces redundant calculations by adopting an incremental approach in the calculation of SAD scores. In addition, it utilizes the parallel SIMD instructions in order to further reduce processing time. The system currently operates at 1 fps (max disparity range at 32 pixels with a 15 x 15 image neighborhood as shown in Fig. 6) and this is still quite far from the reported frame rates in [14] because color images are used instead of grayscale images, MMX intrinsics in Visual Studio are used instead of the original assembly language instruction set and additional steps are required to unwarp and warp image coordinates for catadioptric images. Nevertheless, since this algorithm is highly parallelizable, a new algorithm which utilizes the power of the GPU on a laptop is currently under development. This algorithm will make use of the Compute Unified Device Architecture (CUDA) SDK and Toolkit

developed by Nvidia and processing large images will not be an issue anymore by tapping into the many-core parallel processing power of GPUs.
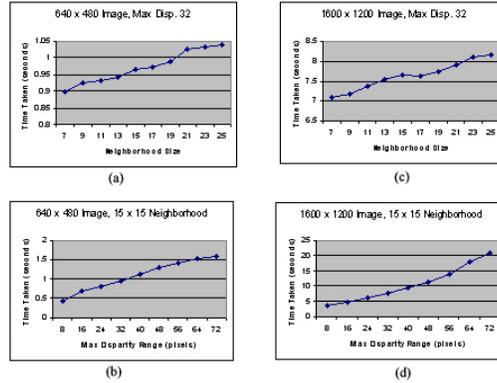
Fig. 6. (a)Time Taken vs. Max Neighborhood Size (640 x 480), (b)Time Taken vs. Max Disparity Range (640 x 480), (c)Time Taken vs. Neighborhood Size (1600 x 1200) and (d)Time Taken vs. Max Disparity Range (1600 x 1200)

Experimental images are collected in an outdoor environment with many natural features such as the two images starting from the top left hand corner in Fig. 7. The disparity map (lighter pixels indicate scene points are closer to the robot) produced by the stereo algorithm is shown and the estimated 3D positions of the stereo correspondences are displayed.
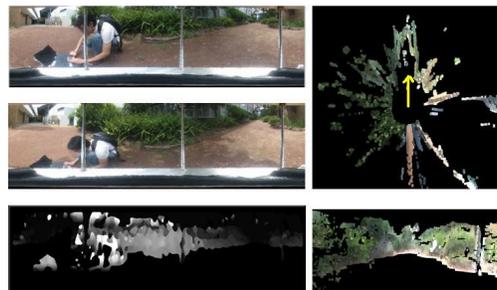
Fig. 7. (Right) Unwarped images and resultant disparity map with an effective baseline of 30cm, (Top Left) plan view of 3D point clouds and (Bottom Left) directional view of the 3D point clouds assuming a camera is placed in middle of the environment facing the direction of the arrow which can be found in the plan view of the 3D point clouds (best viewed in color)

The error of the estimated 3D position of scene points can be found in Fig. 8. Stereo correspondences are selected manually in the image and the estimated depth is compared against ground truth. This process is repeated three times such that the average error can be computed. As illustrated, the accuracy of the estimated 3D positions are affected by both the accuracy of the manually selected stereo correspondences and the information provided in the lookup table. This graph is merely a rough approximation of the accuracy of the omnidirectional system at an effective baseline of 30cm and can be further improved by providing more accurate geometrical values of the system and points on the calibration grid during the calibration process, and manually selected stereo correspondences with subpixel accuracy.

In order to further reduce ambiguity in stereo matches and improve the estimated 3D positions of scene points, plans have been made since early stages of the project to utilize the camera elevation device to research and develop a suitable variable multibaseline catadioptric stereovision algorithm. Although multibaseline stereovision algorithms suffer from high computation requirements but the developments made in GPUs as a general programming device will remove such constraints.
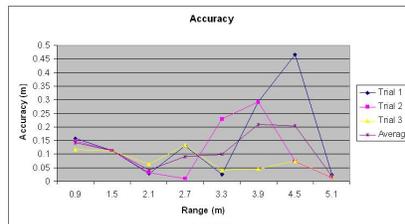
Fig. 8. Accuracy of estimated 3D positions of scene points for stereo pairs with an effective baseline of 30cm

With this information, the robot is able to perceive the environment and initial path planning can be performed and refined as the robot explores the environment by utilizing Distance Transform [15]. This method can be easily extended to 3 dimensions and credibility, tolerance, accuracy and risk factors can be taken into consideration. The ultimate goal of this project is to develop a dynamic expert system which is able to determine the optimum path, not in terms of distance but in terms of a mixture of the previously stated factors, and as well as the best effective baseline to use and the number of stereo images to combine with at a given environment.

6   Conclusion

In this paper, an omnidirectional vision system intended for outdoor mobile robot navigation and path planning is proposed. A new calibration technique for

catadioptric systems with equiangular mirrors was described and a new camera elevation device is designed to support the research and development of a variable multibaseline catadioptric stereovision system. With the promising initial results of the catadioptric stereovision system and potential future research directions, we are one step closer to the realization of a fully autonomous vision-based outdoor mobile robot.

References

1. Benosman, R. and Kang, S. B. Panoramic Vision: Sensors, Theory and Applications. Springer-Verlag, New York, 2001.
2. Baker, S. and Nayar, S. K. A Theory of Single Viewpoint Catadioptric Image Formation. *International Journal of Computer Vision*, 35(2), 175-196, 1999.
3. Svoboda, T. and Pajdla, T. Epipolar Geometry for Central Catadioptric Cameras. *International Journal of Computer Vision*, 49(1), 23-37, 2002.
4. Swaminathan, R., Grossberg, M. D. and Nayar, S. K. Caustics of Catadioptric Cameras. *Eight IEEE Conference on Computer Vision*, 2-9, 2001.
5. Gonçalves, N. and Araújo, H. Projection Model, 3D Reconstruction and Rigid Motion Estimation from Non Central Catadioptric Images. In *2nd Symposium on 3D Data Processing, Visualization and Transmission*, 2004.
6. Mičušík, B. and Pajdla, T. Autocalibration and 3D Reconstruction with Non-central Catadioptric Cameras. *Proc. of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1, 58-65, 2004.
7. Chahl, J. S. and Srinivasan, M. K. Reflective Surfaces for Panoramic Imaging. *Applied Optics*, 36(31), 8275-8285, 1997.
8. Ollis, M., Herman, H. and Singh, S. Analysis and Design of Panoramic Stereo Vision using Equiangular Pixel Cameras. CMU-RI-TR-99-04, Robotics Institute, Carnegie Mellon University.
9. Wheatstone, C. Contributions to the Physiology of Vision. Part the First. On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision. *Philosophical Transactions of the Royal Society of London*, 128, 371-394, 1838.
10. Agrawal, M. and Konolige, K. Real-time Localization in Outdoor Environments using Stereo Vision and Inexpensive GPS. In *The 18th International Conference on Pattern Recognition*, 2006.
11. Little, J., Se, S. and Lowe, D. Vision Based Mobile Robot Navigation and Mapping using Scale Invariant Features. *Proceedings of the IEEE International Conference on Robotics and Automation*, 2051-2058, 2001.
12. Scharstein, D., Szeliski, R. and Zabih, R. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, 47(1/2/3), 7-42, 2002.
13. Fua, P. Combining Stereo and Monocular Information to Compute Dense Depth Maps that Preserve Depth Discontinuities. In *12th International Joint Conference on Artificial Intelligence*, 1292-1298, 1991.
14. Stefano, L. D., Marchionni, M., Mattocia, S. and Neri, G. A Fast Area-Based Stereo Matching Algorithm. *Image and Vision Computing*, 22(12), 983-1005, 2004.
15. Jarvis, R. Distance Transform Based Path Planning for Robot Navigation. in *Recent Trends in Mobile Robots*,Y. F. Zheng, Ed. Singapore:World Scientific, 1993, vol. 11, Robotics and Intelligent Systems, ch. 1.