Depth Perception with a Single Camera

Jonathan R. Seal[1], Donald G. Bailey[2], Gourab Sen Gupta[2]

[1] Institute of Technology and Engineering, [2] Institute of Information Sciences and Technology,
Massey University, Palmerston North, New Zealand
jonoseal12@hotmail.com, d.g.bailey@massey.ac.nz, g.sengupta@massey.ac.nz

Abstract

A catadioptric stereo system has been developed for depth perception. The system uses two sets of planar mirrors to create two virtual cameras. A design procedure is presented with the aim of building a compact assembly. This has resulted in an inexpensive and compact system that achieves a depth resolution of 5.8 cm at a working distance of 2 m.

Keywords: Stereo imaging, planar mirrors, catadioptric system

1 Introduction

When an object moves through space it is restricted in its movement by other objects in its way. For efficient movement through space the object has to know the distance to obstacles and the boundaries of its 'world'. Obtaining the distance to objects, or depth information, for artificial systems has been the subject of many studies, and resulted in many varied methods being proposed [1-13]. Depth perception is the ability to estimate the distance to other objects in an environment to a known accuracy.

Stereo imaging has been suggested as an answer to the problem of depth perception for mobile robotics [1, 9]. Stereo imaging uses multiple images of the same scene taken from different camera locations. The multiple images are related in such a way as to provide disparity. Disparity is defined as the relative movement of an object between two or more views. In a stereo imaging system, the cameras are spatially separated, resulting in the disparity being a function of depth. The disparity is found by matching corresponding points in the input images as illustrated in figures 1 and 2.

Objects closer to the camera have a greater disparity of movement between two images, and this is used to calculate the distance to the objects.

2 Review of Stereo Imaging

Some stereo methods that have been put forward are: the two camera conventional stereo method, which
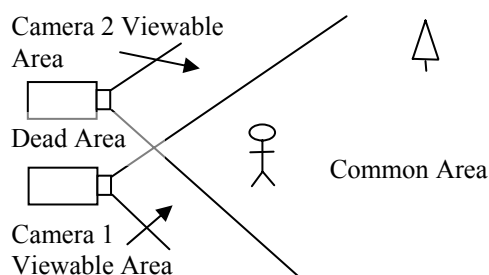

Figure 1: Two camera stereo system.

places multiple cameras on a common axis focused on the same scene with a known baseline; a single camera panning across the scene of interest taking multiple images through its arc; a single camera moving through space and taking multiple images as it moves; and catadioptric stereo which is a single camera using mirrors and lenses to focus on a scene and produce multiple images on the same sensor.

2.1 Conventional Parallel Stereo

Conventional stereo vision is usually achieved with two cameras that are mounted in a known relationship to each other and are synchronised to take images at the same instant.
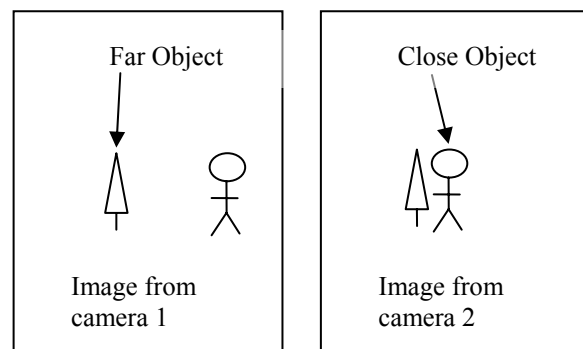

Figure 2: Close and far objects seen through a parallel stereo system with two cameras.

The cameras are usually mounted in parallel (as in Figure 1) as this simplifies the geometry. With parallel cameras, an object point appearing in both images will be offset horizontally between the two images, with the offset, or disparity being inversely proportional to the range of the object point. By ensuring that the cameras are parallel, any perspective distortion will be common to the two images, and does not need to be considered in matching points in one image with those in the other. The main distortion in the images is only from the distortions of the camera lenses, so there is less complex post-processing required.

This is often the method of choice as it is relatively easy to set up. A lot of research has been done on this particular form of implementation [1-3].
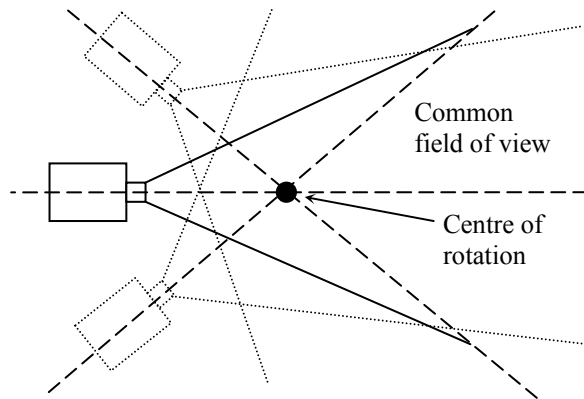
2.2 Stereo with camera panning

Figure 3: Panoramic camera system.

Rather than use 2 cameras, an alternative approach is to use a single camera and pan the camera to obtain images from different views. In Figure 3, the camera is rotated about a point in front of the camera. With this system the line of sight converges, which results in a larger field of view common to both the images than with the parallel stereo [4, 5].

When the line of sight of the cameras is not parallel, perspective distortion must be taken into account when matching points in one image with those in the other image(s).

A disadvantage of panned stereo is that the images must be taken at different times. This will limit the speed with which the distances may be sampled.

2.3 Stereo from a moving platform

A related method, as shown in Figure 4, is a system that works on the camera being fixed and taking images as it travels through its environment. This system also works on multiple images, and uses knowledge of the camera motion between frames to estimate the range to objects.
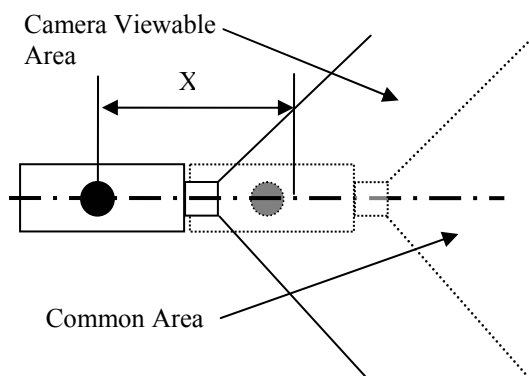

Figure 4: Single camera moving through known space system.

Complex processing is required to find common pixels in both images, as the camera is usually moving in 3D space, and even if it is only moving in 2D space there are many dynamic distortions that can occur if the camera turns from moving directly ahead. When obtaining stereo images from a moving platform using sequential frames, the range estimation is further complicated if the objects are not stationary.

2.4 Catadioptric Stereo

A catadioptric system uses both lenses and mirrors as focusing elements. For stereo imaging the principle is that the lenses and mirrors are designed to act in a way that produces two images on the same sensor as shown in Figure 5 [6, 7].
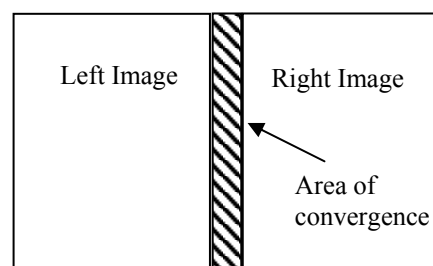

Figure 5: Map of the images on the sensor.

The advantage of catadioptric stereo is that both images are automatically captured at the same time, giving a faster update rate. This method has several drawbacks. The senor is split into two images so that approximately only half the resolution of the original sensor is available for each image. There is also an area of convergence between the two images, which means that some of the sensor area is lost.

Many catadioptric systems have been developed using rounded mirrors [8 - 10] to get uni-polar vision, so as to see all around the robot at once. There are also several other catadioptric configurations using planar mirrors, some of which are discussed in [11 - 13].
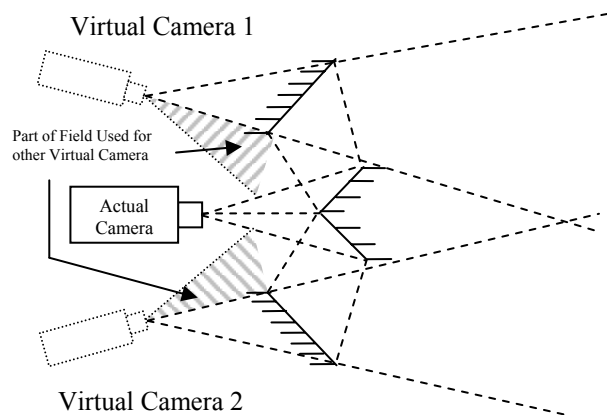
3 System Design

Figure 6: Our catadioptric stereo vision system.

Our goal was to design a compact catadioptric stereo system for a mobile robot that has a field of view of 1 meter at a range of 2 meters. The target depth resolution is 5 cm.

Our design is similar to the work described in [6]. Figure 6 shows our systems layout, as well as the position of the virtual cameras. Two mirrors are used to split the field of view, and another two to reflect the two new fields of view to overlap in the region of interest.

All four mirrors are fixed in a mounting, to which the camera is also mounted. The mounting is machined, and designed in such a way that an intricate assembly process is not required to align the system.

3.1 System Geometry

We used a pinhole model to describe the camera and then we used Microsoft Excel to model and design the mirror system. As there was symmetry in the system design, only half of the mirror assembly had to be designed.

Figure 7 shows the system geometry. The following are the definitions of the symbols used:

ε = The sensor size

f = focal length of the lens

θ = The angle formed by the outside ray from the sensor that passes through the pinhole, where

$$\tan\theta = \frac{\varepsilon}{2 \cdot f} \tag{1}$$

η = the distance of the pinhole from the vertex of the inside mirrors

ϕ = the inside mirror angle

α = outside mirror angle

D = working distance

FOV = field of view at working distance

To give the most compact design, the central ray reflected off the inner and outer mirrors must just clear the outside edge of the inner mirror. To obtain the maximum field of view at the working distance for a particular configuration, the angle of the outside mirror is adjusted so that the inner ray from one view intersects the outer ray of the other view at the working distance.

The field of view is controlled by the angle of view of each virtual image. To a first approximation (and for small θ), the field of view at the working distance is given by:

$$FOV \approx D.\tan\theta \tag{2}$$

Rearranging this gives:

$$\tan\theta \approx \frac{FOV}{D} \tag{3}$$

By equating (1) and (3) we can solve for the focal length needed to achieve the desired field of view:

$$f = \frac{\varepsilon}{2\tan\theta} \approx \frac{\varepsilon D}{2FOV} \tag{4}$$

This allows the lens to be selected to achieve a particular working width for a given sensor size.

When designing a depth perception system it is important to design the system around the required depth accuracy. Let

B = baseline or the distance between the pinholes in the virtual cameras

λ = the width of one pixel on the sensor

The depth resolution d, at the working depth D can be defined as the change in range that results in a change in the disparity of 1 pixel in the image. Figure 8
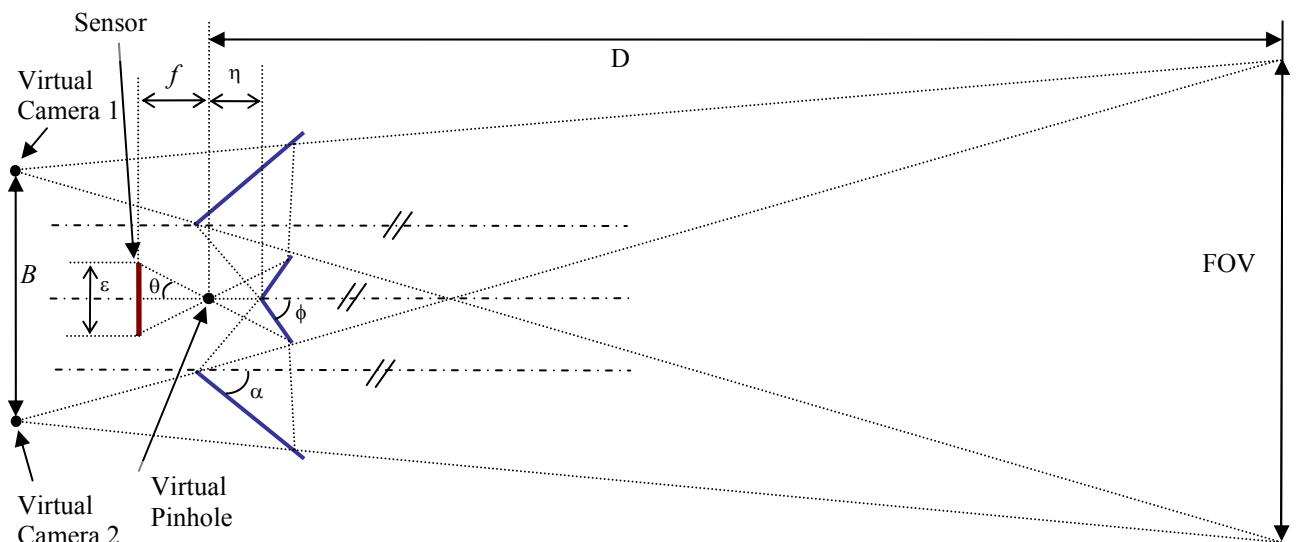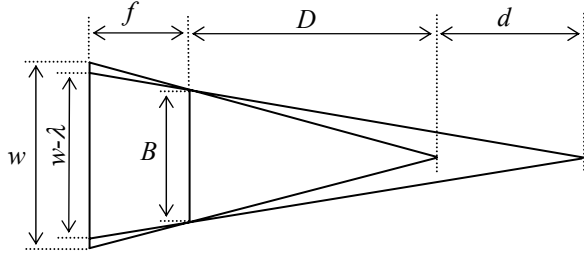

Figure 7: The catadioptric system showing the position of mirrors, pinhole and the CCD.

shows this geometrical arrangement for calculating the depth resolution.

Figure 8: Moving an object point by a distance *d* changes the disparity between the images by *λ*.

From similar triangles

$$\frac{w}{D+f} = \frac{B}{D} \tag{5}$$

and

$$\frac{w-\lambda}{D+d+f} = \frac{B}{D+d} \tag{6}$$

Eliminating *w* from (7) and (8) and rearranging gives the minimum baseline required to achieve a particular depth resolution as

$$B = \frac{\lambda D}{fd}(D-d) \tag{7}$$

or for a given arrangement, the depth resolution is given by

$$d = \frac{\lambda D^2}{Bf - \lambda D} \tag{8}$$

Note that because the two virtual cameras have converging lines of sight, the above analysis is only an approximation. The introduction of perspective distortion will mean that the pixel resolution will depend on the actual location of the target point in the image. Depth resolution may also be improved by using sub-pixel matching techniques [14].

For the geometric arrangements of Figure 6 and 7, the lengths of the mirrors may be found by using the sine rule. The inner mirrors have length:

$$M_{in} = \eta . \frac{\sin\theta}{\sin(\phi-\theta)} \tag{9}$$

and the outer mirrors:

$$M_{out} = \eta . \frac{\sin\theta \sin\phi \sin(4\phi - 2\alpha - \theta)}{\sin(\phi-\theta)\sin(4\phi-2\alpha)\sin(2\phi-\theta-\alpha)} \tag{10}$$

Similarly, the baseline was derived to be:

$$B = 2\sin(2\phi - 2\alpha)\eta \times$$
$$\left(1 + \frac{\sin\theta}{\sin(\phi-\theta)} \cdot \left(\frac{\sin\phi}{\sin(2\phi-2\alpha)} + \frac{\sin(3\phi-2\alpha+\sin\phi)}{\sin(4\phi-2\alpha)}\right)\right) \tag{11}$$

Using a Microsoft Excel spreadsheet allowed us to perform all our calculations in the same document. It also enabled us to see the values of such variables as the height and width of the mirrors and the length of the inside and outside rays. Most importantly Excel allowed us to be able to test many scenarios and get full data for each scenario easily and quickly.

3.2 The Design tradeoffs

The physical size of the system is limited by the application that it is used for. If the mirror assembly could never be designed small enough to be mounted on a medium sized autonomous agent (400mm x 400 mm x 400 mm), then it would never find any application outside of the laboratory.

The major advantage of the single camera stereo system is cost. Only a single camera and frame grabber is required. The use of catadioptric stereo ensures that the two virtual images are captured simultaneously. The main disadvantage of the arrangement shown here is that the virtual cameras are not parallel. This introduces perspective distortion, which cannot be avoided in this configuration. If the two cameras were parallel, there would be no common area of view between the two images.
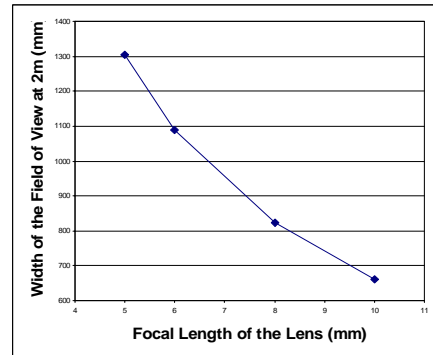

Figure 9: Effect of the lens focal length on the field of view for a 6.4 mm x 4.4 mm sensor with a working depth of 2 meters.

Equations (1) and (2) suggest that the primary factor affecting the field of view is the focal length. This was verified using the spreadsheet, with the relationship shown in figure 9.

A narrow FOV will give improved spatial resolution. Since depth resolution also depends on spatial resolution, it will also improve with a narrow field of view.

The field of view has to be reasonable, because a narrow field of view would restrict the usefulness of the system on an autonomous mobile system, as the robot would not be able to see far enough into its periphery to avoid collision with other mobile systems. Too large a field of view, while not a problem in itself, would reduce both the depth and spatial resolution.

99
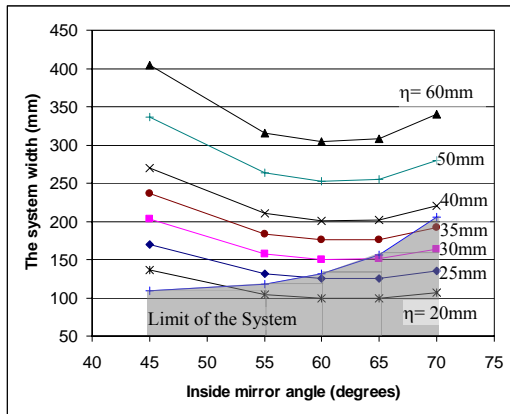
Figure 10: The effect of the distance of the pinhole from the vertex of the inside mirrors, on the system size. The shaded region is where the system is self-occluding.

The size of the lens and mirror system is directly proportional to the distance between the pinhole and the inner mirrors. This results from similar triangles and is reflected in equations (9) and (10). Prior work with this configuration fixed the angle of the inner mirrors to be 45° [6, 7]. Increasing this angle has the potential to improve the compactness of the design because the length of the inner mirror will decrease, allowing a narrower baseline. This effect is clearly illustrated in Figure 10. A mirror angle of approximately 60° gives the most compact design, keeping the other parameters the same.

Angles greater than 45° fold the optical path back behind the inner mirror. An additional constraint under these conditions is that the lens itself does not occlude the objects being viewed. The physical size of the lens will limit the distance from the pinhole to the inner mirrors as illustrated in figure 11.

With reference to figure 11, d is the width of the lens and p is the distance of the pinhole from the front of the lens
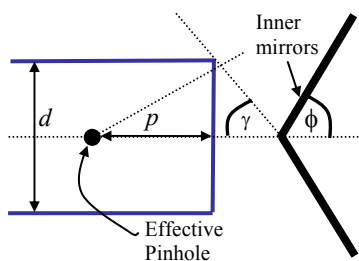
Figure 11: The lens and mirror arrangement to show how the size of the lens limits how close the pinhole can be to the vertex of the inside mirrors.

The central ray is reflected off the mirrors with angle

$$\gamma = 180 - 2\phi \qquad (12)$$

From trigonometry, the system is constrained such that

$$\tan \gamma > \frac{d/2}{\eta - p} \qquad (13)$$

Substituting (12) into (13) and rearranging gives

$$\eta > p - \frac{d/2}{\tan 2\phi} \qquad (14)$$

This constraint has been mapped onto figure 10.

The use of a pinhole camera model becomes less valid as the pinhole to mirror distance is decreased. This appears in the image as a region of convergence, or blur, between the two views. The angle subtended by the join between the mirrors also increases as the lens is moved closer to the mirrors.

4   Implementation

High grade polished aluminium was used for the mirrors because this is less expensive than front silvered glass mirrors. Polished aluminium has the disadvantage of being very easily damaged, and once damaged it can not be very easily repaired.

Figure 12: The completed catadioptric system.

Aluminium was used to build the rest of the system with three 10mm plates being used as supporting structures. The structure can be seen in figure 12.
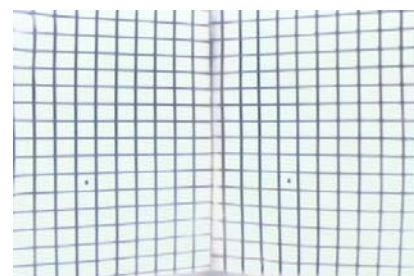
Figure 13: The calibration grid seen through the vision system

Figure 13 shows the image of a 10 cm spaced grid placed 2 m from the camera. This image may be used to calibrate the system to correct for distortions within the system. These arise from three main sources. The most obvious is the perspective distortion in each of the views. This results from the converging line of sight of the two virtual cameras, and is a direct result of the plane at 2 m not being parallel to the image

plane within the virtual cameras. The second source of distortion is barrel lens distortion from the use of a wide angle lens. Magnification in the centre of the image is slightly greater than in the periphery, and this results in the characteristic barrel shape of the vertical lines. The third source of distortion results from minor distortions in the polished aluminium mirrors. These result in apparently random deviations or twists in the lines. The calibration image provides information that is able to characterise and correct these distortions, allowing the true disparity between the two sides to be measured.

5   Summary and Future Work

In this paper we have detailed the design procedure to build a compact catadioptric stereo system for depth perception. The system uses a single camera and planar mirrors which were made of aluminium. The system is inexpensive and is targeted for use on a mobile robot platform. From equation (8), theoretically the system achieves a depth resolution of 5.8 cm at a working distance of 2 m.

The next step is to use the calibration image to characterise the distortions. This calibration will result in a pair of images with no disparity for objects at the working distance. After correcting the distortions, a disparity map may then be produced by matching corresponding points between left and right images. With the distortion removed, standard parallel line of sight stereo algorithms may be used. From the disparity map, a 2½-D depth map may be produced to provide data for robot navigation.

6   Acknowledgements

We would like to thank Ken Mercer and Colin Plaw for their input into the formation and evolution of the concepts of this design. We would also like to thank Leith Baker and Humphrey O'Hagan from the metal and CAM workshop for their dedication to completing the prototype in a timely fashion.

7   References

[1] Hariti, M., Ruichek, Y., Koukam, A., "A fast stereo matching method for real time vehicle front perception with linear cameras", *IEEE Proceedings Intelligent Vehicles Symposium,* pp 247-252 (2003).

[2] Chiou, R.N., Chen, C.H., Hung, K.C., Lee, J.Y., "The optimal camera geometry and performance analysis of a trinocular vision system", *IEEE Transactions on Systems, Man and Cybernetics,* 25:8, pp 1207-1220 (1995)

[3] Venter, M.A.H., van Schalkwyk, J.J.D., "Stereo imaging in low bitrate video coding", *Southern African Conference Proceedings on Communications and Signal Processing. (COMSIG),* pp 115-118 (1989).

[4] Fabrizio1, J., Tarel, J.P., Benosman, R., "Calibration of Panoramic Catadioptric Sensors Made Easier", *Proceedings of Third Workshop on Vision,* pp 45-52 (2002).

[5] Huang, F., Wei, S.K., Klette, R., Gimelfarb, G., "Cylindrical Panoramic Cameras – From Basic Design to Applications", *Proc. Image and Vision Computing New Zealand conference,* pp 101-106 (2002).

[6] Inaba, M., Hara, T., Inoue, H., "A Stereo Viewer Based on a Single Camera with View-Control Mechanism", *Proc. International Conference Robots and Systems,* pp 1857-1864 (1993).

[7] Mathieu, H., Devernay, F., "Systeme de Miroirs pour la Stereoscopie", *Technical Report 0172, INRIA Sophia-Antipolis*, 1995.

[8] Derrien, S., Konolige, K., "Approximating a Single Viewpoint in Panoramic Imaging Devices", *Proceedings of the IEEE International Conference on Robotics & Automation,* pp 3931-3938 (2000).

[9] Nayar, S.K., "Catadioptric omnidirectional camera", *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* pp 482-488 (1997)

[10] Fiala, M., Basu, A., "Feature extraction and calibration for stereo reconstruction using non-SVP optics in a panoramic stereo-vision sensor", *Proceedings of Third Workshop on Omnidirectional Vision,* pp 79-86 (2002).

[11] Gluckman, J., Nayar, S.K., "Rectified Catadioptric Stereo Sensors", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 24:2, pp 224-236 (2002)

[12] Gluckman, J., Nayar, S.K., "Planar catadioptric stereo: geometry and calibration" *IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* Volume 1, pp 22-28 (1999).

[13] Gluckman J., Nayar S.K., "Catadioptric stereo using planar mirrors", *International Journal of Computer Vision,* 44, pp 65-79 (2001).

[14] Bailey D. G., "Sub-pixel estimation of local extrema", *Proceeding of Image and Vision Computing New Zealand,* pp 414-419 (2003).