Omnidirectional Projections with a Cone Mirror and Single Mirror Stereo

Chris Burbridge, Ulrich Nehmzow, Joan Condell

Omnidirectional Projections with a Cone Mirror and Single Mirror Stereo

Chris Burbridge, Ulrich Nehmzow, and Joan Condell

School of Computing and Intelligent Systems, University of Ulster, Magee Campus, Londonderry, BT48 7JL, Northern Irland
cburbridge@ieee.org u.nehmzow@ulster.ac.uk j.condell@ulster.ac.uk

Abstract. In this paper we present forward and backward projection formulae for catadioptric omnidirectional vision using an arbitrarily located cone mirror and perspective camera. For the back projection a quasi closed-form solution is derived, while for the forward projection an itterative method is used. We then show that these formulae can be used to achieve stereo vision using only a single mirror, without strict alignment requirements between the cameras and the mirror.

1 Introduction

There are many different methods of achieving omnidirectional vision. A pyramid of mirrors are used in [1], with multiple cameras pointing at the pyramid. This configuration offers high resolution and the possibility of a single view point, but is not isotropic, and the registration and the physical arrangement of the cameras can be difficult. Rotating cameras and mirrors were used in [2] and [3]. However, difficulties were encountered with the registration and the motion delay of the camera. Wide angle and fisheye lenses have been used in a purely dioptric setup in [4] and [5], and methods of rectifying the distortion of the images caused by these lenses have been proposed [6].

There are several types of rotationally symmetric mirror that can be used in a catadioptric setup. Parabolic, hyperbolic and ellipsoidal mirrors can be used to achieve a single effective viewpoint – something that is usually considered important for many applications. However, using a cone mirror is preferable as the camera resolution is spread over a usefull area for robotics applications [7], unlike many single view point sensors such as fish-eye lenses which waste resolution in the central part of the image looking at the sky or ground. Cone mirrors have been used by a number of people ([8], [9], [10], [11], [12], [13]), but in all cases either the camera is assumed to be perfectly aligned with the cone axis, or exact geometric information is not required.

The ability to measure the depth to image features is important and useful in many applications of omnidirectional vision, for example robot navigation. Methods for achieving stereo omnidirectional vision are now abundant, but catadioptric setups often require strict alignment of the mirrors and cameras and dioptric methods need careful lens calibration to remove distortions.

Catadioptric stereo is usually obtained by placing two cameras and mirrors co-axially as in [14] [15] and [16]. While this has shown to be an effective method both in simulation experiments ([17]) and with well manufactured sensors ([18]), when the sensors are poorly aligned the system will fail.

Configurations that require only a single camera have been proposed in [19] and [20]. In [19] two mirrors are placed co-axially, with the a hole in the centre of the lower mirror for the camera to look through. The lower mirror is then seen reflected in the upper mirror, providing two views of the environment in one image. In [20] a single camera is mounted looking down at multiple side by side spherical mirrors. Using only a single camera for stereo is advantageous because the optical response will be the same for both views. However, both of these methods require carefull alignment, and the already low resolution of a single camera has to be shared between multiple views.

Stereo with a single mirror is a desirable goal in many applications, where the mirrors are either too expensive, or too heavy to carry two of. Stereo with a single cone mirror was proposed in [10]. This method made use of a beam splitter to allow the cameras to be both aligned with the cone axis. Again, perfect alignment is required for the stereo to be achieved.

Forward projection formula for projection through a missaligned cone have been proposed in [21]. However, the formula proposed will only work if the vertex of the cone, the world point, the camera and the normal to the cone are all in the same plane. In this paper we present forward and backward projection formulae for catadioptric omnidirectional vision using an arbitrarily placed cone mirror and perspective camera. We then show that these formulae can be used to achieve stereo vision using only a single mirror, without strict alignment requirements between the cameras and the mirror.

2 Back Projection

The back projection of a pixel provides the ray line in 3 dimensional space on which the point imaged by the pixel must lie. The ray line is described in vector form as $(\mathbf{D}, \hat{\mathbf{d}})$ where $\mathbf{D}$ is the start point of the ray and $\hat{\mathbf{d}}$ is a unit vector giving the direction. Tracing the ray backwards from the camera can be split into three stages: (i) determining the direction of the ray as it leaves the camera, (ii) determining the point on the surface of the mirror at which it was reflected, and (iii) determing the direction of the ray from the point on the mirror to the imaged point. These stages are found as follows.

2.1 The ray leaving the camera

The direction of the reversed ray leaving the camera can be specified with the standard pinhole camera model using the intrinsic parameters of the camera, namely the focal length f, scale factor s and the coordinates of the principal point p_x, p_y.

In a camera centred coordinate system with the camera looking along the negative z axis and the horizontal image direction parallel to the x axis as shown in Figure 1, the direction of a ray leaving the camera associated with an image location (u, v) is

$$\mathbf{d} = \begin{pmatrix} p_x - u \\ p_y - v \\ f/s \end{pmatrix}. \tag{1}$$

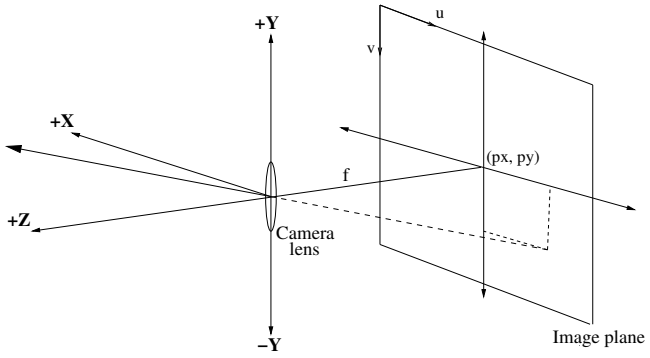

Fig. 1. The perspective projection camera model.

We convert the ray from camera centred coordinates to cone centred coordinates, as shown in Figure 2. The ray transforms to $(\mathbf{C}, \hat{\mathbf{c}})$, where

$$\hat{\mathbf{c}} = R\hat{\mathbf{d}}. \tag{2}$$

R is a rotation matrix built up of the Euler angles for the rotation of the camera, and $\mathbf{C}$ is a vector giving the translation between cone and camera coordinates.

2.2 Intersection with the cone

The point of intersection $\mathbf{I}$ between the ray and the camera can be found algebraically by equating the formula for the cone with the formula for the ray:

$$\begin{pmatrix} I_x \\ I_y \\ I_z \end{pmatrix} = \mathbf{I} = \mathbf{C} + p\hat{\mathbf{c}}, \tag{3}$$

$$I_z + \frac{\sqrt{I_x^2 + I_y^2}h}{r} = 0. \tag{4}$$

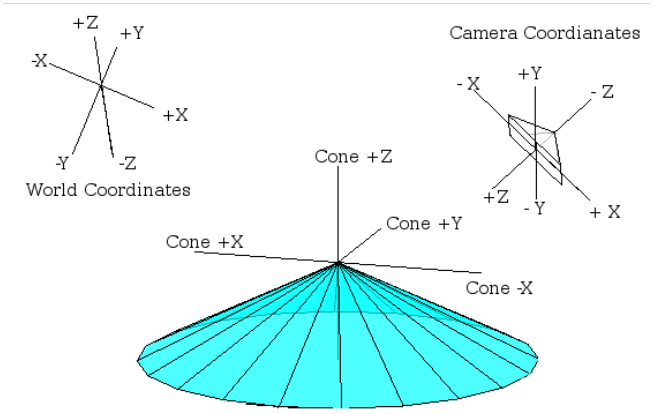

Fig. 2. The three coordinate systems: world, cone and camera. Each coordinate system is unique, with a rotation and translation between systems.

Substituting (3) into (4) and solving for p gives two solutions:

$$\begin{aligned} p = -((-A_x - A_y)a^2 + A_z - (2A_x a^2 A_y - 2A_x A_z \\ - 2A_y A_z + c_x^2 C_z^2 - c_x^2 a^2 C_y^2 - c_y^2 a^2 C_x^2 \\ + c_y^2 C_z^2 + c_z^2 C_x^2 + c_z^2 C_y^2)^{\frac{1}{2}} a)/(-c_x^2 a^2 - c_y^2 a^2 + c_z^2); \end{aligned} \tag{5}$$

$$\begin{aligned} \text{and } p = -((-A_x - A_y)a^2 + A_z + (2A_x a^2 A_y - 2A_x A_z \\ - 2A_y A_z + c_x^2 C_z^2 - c_x^2 a^2 C_y^2 - c_y^2 a^2 C_x^2 \\ + c_y^2 C_z^2 + c_z^2 C_x^2 + c_z^2 C_y^2)^{\frac{1}{2}} a)/(-c_x^2 a^2 - c_y^2 a^2 + c_z^2); \end{aligned} \tag{6}$$

where $a = \frac{h}{r}$ is the ratio between the height h and radius r of the cone, $A_x = C_x \hat{c}_x$, $A_y = C_y \hat{c}_y$ and $A_y = C_y \hat{c}_y$. The correct value of p is selected as the smallest of the solutions that satisfy $C_z + pc_z < 0$.

2.3 The direction of the reflected ray

We have found that the back projected ray intersects the mirror at $\mathbf{I}$ with a direction vector $\hat{\mathbf{c}}$. We now assume that that the mirror is a perfectly specular reflective surface and that there are no manufacturing imperfections in the shape of the cone. The reflected ray direction can then be found as

$$\hat{\mathbf{r}} = \hat{\mathbf{c}} - 2(\hat{\mathbf{c}} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}} \tag{7}$$

where $\hat{\mathbf{n}}$ is the normalised vector normal to the cone at $\mathbf{I}$ found by differentiating the formula for the cone (4) at I.

$$\begin{aligned}\mathbf{n} = \Delta f &= \begin{pmatrix} \frac{\partial f}{\partial I_x} \\ \frac{\partial f}{\partial I_y} \\ \frac{\partial f}{\partial I_z} \end{pmatrix} \\ &= \begin{pmatrix} hI_x/((I_x^2+I_y^2)^{\frac{1}{2}}r) \\ hI_y/((I_x^2+I_y^2)^{\frac{1}{2}}r) \\ 1 \end{pmatrix}. \end{aligned} \tag{8}$$

Multiplying by $\sqrt{I_x^2+I_y^2}$ and observing that $\frac{h}{r}$ is the same as $\frac{I_z}{(I_x^2+I_y^2)^{-1/2}}$ allows (8) to be simplified to

$$\mathbf{n} = \begin{pmatrix} I_xI_z \\ I_yI_z \\ I_x^2I_y^2 \end{pmatrix}. \tag{9}$$

The resulting ray line that the point imaged at (u, v) must lie on is $(\mathbf{I}, \hat{\mathbf{r}})$.

2.4 Model Validation

In order to test the model two artificial images were generated using PovRay, a ray tracing package that accurately simulates vision by tracing rays through a scene. As shown in Figure 3, in one image the camera was placed so that it was nearly aligned with the mirror axis and in the other it was placed looking at the side of the mirror. Image points corresponding to known world points were then human selected to the nearest pixel, and the back projection was performed to find the corresponding reflected ray. For each selected image point, the distance from the known world point to the calculated ray was found.

In both images the average distance to the selected points was 2.4 meters. For image 1 where the camera was almost aligned with the axis the average distance between the world point and the calculated ray was 8.3mm, and for image 2 the average distance between world point and ray was 5.7mm. The higher level of accuracy for the second image can be attributed to the fact that the same resolution is covering a smaller region, reducing the effect of the error in the hand selected image coordinates.

3 Forward Projection

The aim of the forward projection model is to map a point from world coordinates $\mathbf{P} = \begin{pmatrix} x & y & z \end{pmatrix}^T$ to the pixel coordinates (u, v) at which it appears in the image. This involves two steps: (i) finding the point on the surface of the mirror that reflects the point into the camera (the reflection point), and (ii) projecting this point into the camera. Finding the reflection point can be achieved by placing constraints upon the location, and then solving these constraints as a system of equations.

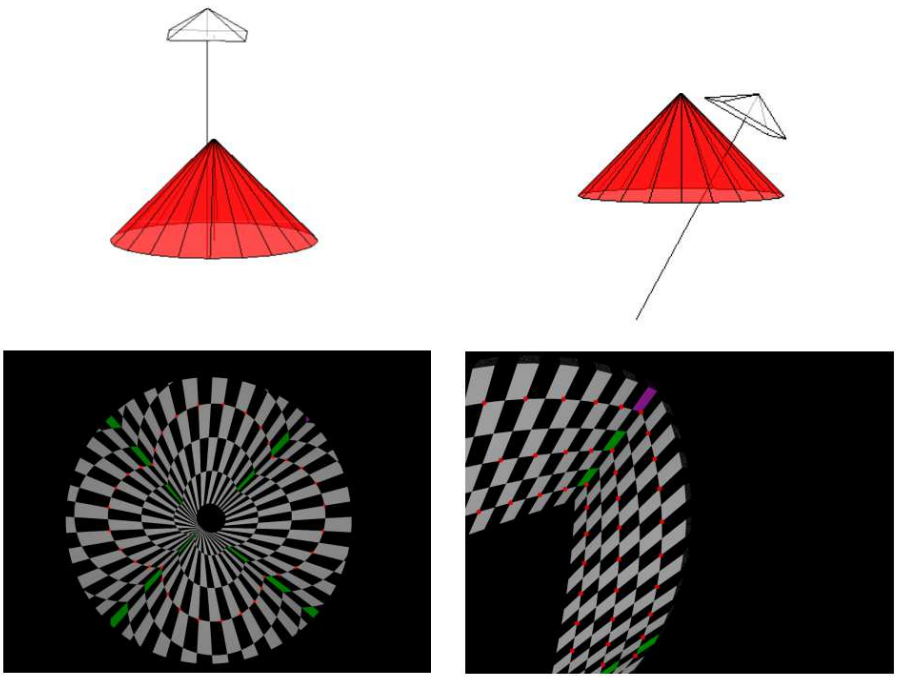

Fig. 3. The two setups and generated images for testing the projection models. On the left the camera is almost aligned with the axis of the mirror, while on the right the camera looks at the mirror from the side.

3.1 The reflection point

Assuming that the world point has been transformed into cone centred coordinates with the rotation and translation between cone and world known, we have the following constraints upon the reflection point $\mathbf{R} = \begin{pmatrix} R_x & R_y & R_z \end{pmatrix}^T$:

1. The point must be on the surface of the cone:

$$R_z + \sqrt{R_x^2 + R_y^2}\frac{h}{r} = 0. \tag{10}$$

2. The angle between the normal to the cone at the reflection point and the vector from the world point to the reflection point must equal the angle between the normal and vector from the camera to the reflection point:

$$\frac{\mathbf{C} - \mathbf{R}}{\|\mathbf{C} - \mathbf{R}\|} \cdot \mathbf{n} - \frac{\mathbf{P} - \mathbf{R}}{\|\mathbf{P} - \mathbf{R}\|} \cdot \mathbf{n} = 0 \tag{11}$$

expanding to

$$\frac{((C_x - R_x)R_zR_x + (C_y - R_y)R_yR_z + (C_z - R_z)(-R_x^2 - R_y^2))^2}{(C_x - R_x)^2 + (C_y - R_y)^2 + (C_z - R_z)^2} - \frac{((x - R_x)R_zR_x + (y - R_y)R_yR_z + (z - R_z)(-R_x^2 - R_y^2))^2}{(x - R_x)^2 + (y - R_y)^2 + (z - R_z)^2} = 0. \quad (12)$$

3. The normal to the cone at the reflection point must be in the same plane as the world point, the camera, and the reflection point:

$$((\mathbf{P} - \mathbf{C}) \times (\mathbf{P} - \mathbf{R})) \cdot \mathbf{n} = 0 \quad (13)$$

expanding to

$$\begin{aligned} &((y - C_y)(z - R_z) - (z - C_z)(y - R_y))\ R_zR_x + \\ &((z - C_z)(x - R_x) - (x - C_x)(z - R_z))\ R_zR_y + \\ &((x - C_x)(y - R_y) - (y - C_y)(x - R_x))\ (-R_x^2 - R_y^2) = 0. \end{aligned} \quad (14)$$

A direct analytical solution for $\mathbf{R}$ cannot be obtained from the above constraints because equation (12) is highly non-linear, containing sixth powers. Figure 4 shows the complexity of these constraints graphically.

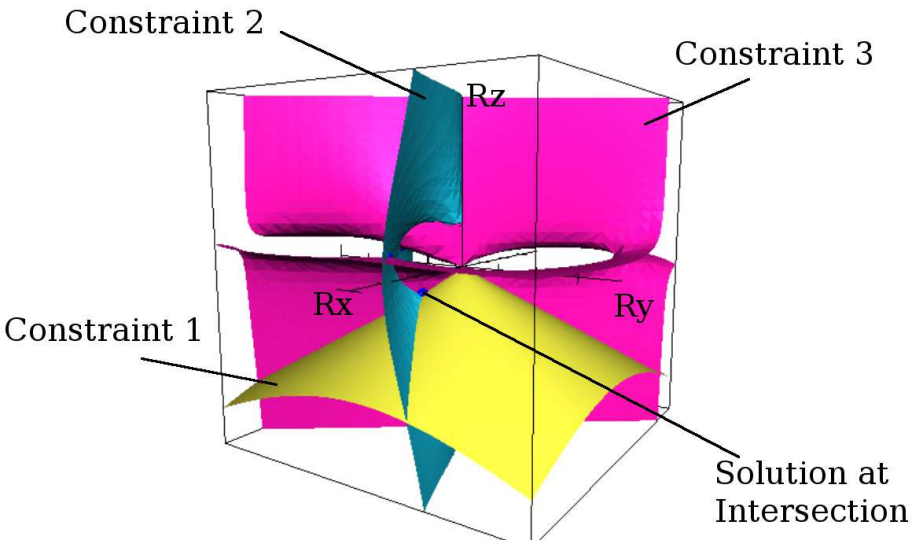

Fig. 4. A graphical representation of the forward projection constraints (10), (12) and (14). All parameters other than $\mathbf{R}$ are fixed.

An iterative solution A solution can be obtained iteratively using Newton's method:

$$\mathbf{R_{n+1}} = \mathbf{R_n} - \Delta F(\mathbf{R_n})^{-1} F(\mathbf{R_n}) \tag{15}$$

where $F = 0$ is a multidimensional function comprising of equations (10), (12) and (14), and ΔF is the Jacobian of F with respect to the reflection point $\mathbf{R_n}$ at iteration n. With a good initial estimate of $\mathbf{R_n}$ few iterations are generally required, however, when a poor initial estimate is used convergence is not guaranteed. Adaptations of Newton's method such as the Newton-Armijo method which addresses this issue by introducing a step size and the Newton-GMRES iteration which offers better global convergence are discussed in [22].

When the camera is nearly aligned with the axis in the usually required manner, the projection equations detailed by Spacek [14] can be used to provide an initial estimate. When the camera is intentionally set off axis as is the case in section 4 a method of searching for an initial point is required.

3.2 Projection into the camera

The reflection point $\mathbf{R}$ in cone centred coordinates can then be projected into the camera using the pinhole camera model:

$$\begin{bmatrix} Su \\ Sv \\ S \end{bmatrix} = M \begin{bmatrix} R_x \\ R_y \\ R_z \\ 1 \end{bmatrix} \tag{16}$$

where S is the homogeneous coordinate scale factor and M is the 3×4 camera projection matrix comprising of both the intrinsic and extrinsic parameters of the camera as in (1) and (2):

$$M = \begin{bmatrix} sf & 0 & p_x & 0 \\ 0 & -sf & p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R^{-1} & -\mathbf{C} \\ \mathbf{0} & 1 \end{bmatrix}. \tag{17}$$

3.3 Model Validation

The forward projection method was tested on the same images as in section 2.4, using the human selected points as the initial starting value for the iteration. In both cases, the average number of iterations was 2, with the maximum being 4. After performing the forward projection, the backward projection was re-applied. The distance between the calculated ray and the world point for all points in both images was then found to be 0, indicating that both models are correct.

4 Stereo

It is now possible to perform stereo vision using a single mirror with two off-axis cameras by using the projection equations described in the previous sections. The equations can be used with any arrangement of mirror and cameras, but in order to perform stereo a common region visible in both images is required. From the two views, common features are backprojected to find the ray line they lie on. As in classical stereo, the reconstructed point is the mid-point of the shortest line between the projection rays.

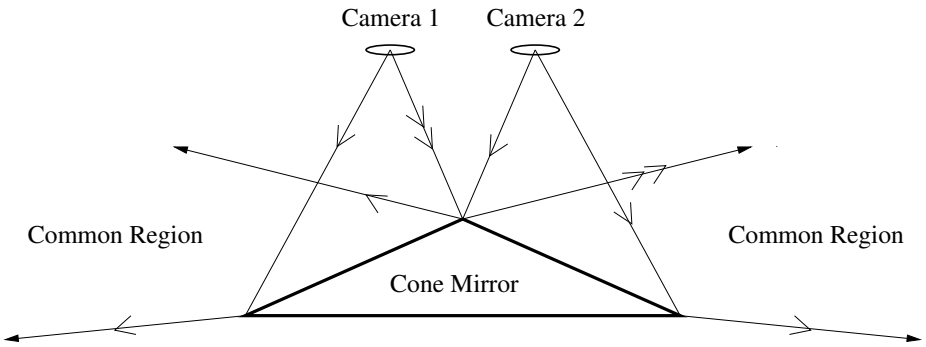

Fig. 5. The single mirror stereo arrangement tested. Note that in a 2D sketch the common region is covered by a very small image area, but in 3D the common region is more noticeable as can be seen in Figure 6.

We tested the arrangement shown in Figure 5 using PovRay. Two views were generated (see Figure 6), and ten common features were human selected to the nearest pixel. The back projection was then carried out and the point was reconstructed. The average distance between the reconstructed point and the ground truth was found to be 193 mm, with a maximum reconstruction error of 367 mm. The average percentage error was 7.5% of the range to the point.

5 Conclusion

In this paper we have shown that it is not necessary to make sure that a camera is perfectly aligned with the mirror axis in order to make use of a cone mirror for omnidirectional stereo. We have derived simple equations for the forward and backward projections for the general case of an arbitrarily placed camera, and have shown in simulation that by using these equations only one mirror is required for stereo.

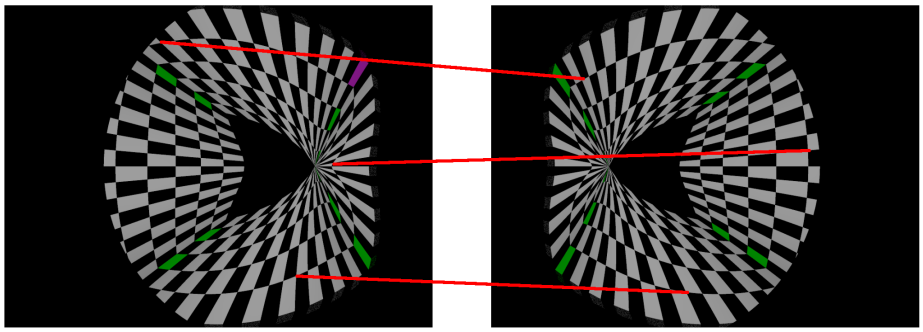

Fig. 6. Two test images generated for testing stereo. Three correspondences are marked.

References

1. Tan, K., Hua, H., Ahuja, N.: Multiview panoramic cameras using mirror pyramids. IEEE Trans. Pattern Anal. Mach. Intell. **26**(7) (2004) 941–946
2. Kang, S.B., Szeliski, R.: 3-d scene data recovery using omnidirectional multibaseline stereo. In: CVPR '96: Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition (CVPR '96), Washington, DC, USA, IEEE Computer Society (1996) 364
3. Ishiguro, H., Yamamoto, M., Tsuji, S.: Omni-directional stereo. IEEE Trans. Pattern Anal. Mach. Intell. **14**(2) (1992) 257–262
4. Swaminathan, R., Nayar, S.K.: Nonmetric calibration of wide-angle lenses and polycameras. IEEE Trans. Pattern Anal. Mach. Intell. **22**(10) (2000) 1172–1178
5. Shah, S., Aggarwal, J.K.: Mobile robot navigation and scene modeling using stereo fish-eye lens system. Mach. Vision Appl. **10**(4) (1997) 159–173
6. Micusik, B., Pajdla, T.: Structure from motion with wide circular field of view cameras. Transactions on Patter Analysis and Machine Intelligence **28**(7) (July 2006) 1135–1149
7. Lin, S.S., Bajcsy, R.: High resolution catadioptric omni-directional stereo sensor for robot vision. In: ICRA, IEEE (2003) 1694–1699
8. Cauchois, C., Brassart, E., Pegard, C., Clerentin, A.: Technique for calibrating an omnidirectional sensor. In: Internation Conference on Intelligent Robots and Systems, 1999. Volume 1. (1999) 166–171
9. Lin, S.S., Bajcsy, R.: True single view point cone mirror omni-directional catadioptric system. In: ICCV. (2001) 102–107
10. Lin, S.S., Bajcsy, R.: Single cone mirror omni-directional stereo. Technical Report MS-CIS-01-03, University of Pennsylvania (2001)
11. Yagi, Y., Kawato, S.: Panoramic scene analysis with conic projection. In: IEEE/RSJ Conference on Intelligent Robots and Systems. (1990)
12. Nayar, S.K., Baker, S.: Catadioptric image formation. In: Proceedings of the 1997 DARPA Image Understanding Workshop, New Orleans, Louisiana (1997) 1431–1347
13. Francis, G., Spacek, L.: Linux robot with omnidirectional vision. In: Proceedings of TAROS 2006. (2006) 56–63

14. Spacek, L.: A catadioptric sensor with multiple viewpoints. Robotics and Autonomous Systems **51**(1) (2005) 3–15
15. Gluckman, J., Nayar, S., Thoresz, K.: Real-time omnidirectional and panoramic stereo. In: DARPA Image Understanding Workshop (IUW). (Nov 1998) 299–303
16. Negishi, Y., Miura, J., Shirai, Y.: Calibration of omnidirectional stereo for mobile robots. In: IEEE International Conference on Intelligent Robots and Systems. Volume 3. (October 2004) 2600–2605
17. Burbridge, C., Spacek, L.: Omnidirectional vision simulation and robot localisation. In: Proceedings of TAROS 2006. (2006) 32 – 39
18. Sogo, T., Ishiguro, H., Trivedi, M.M.: Real-time target localization and tracking by n-ocular stereo. In: OMNIVIS '00: Proceedings of the IEEE Workshop on Omnidirectional Vision, Washington, DC, USA, IEEE Computer Society (2000) 153
19. Gijeong Jang, Sungho Kim, I.K.: Single-camera panoramic stereo system with single-viewpoint optics. Optics Letters **31**(1) (2006) 41–43
20. Mouaddib, E.M., Dequen, G., Devendeville, L.: A new omnidirectional stereovision sensor. Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on (Oct. 2007) 1–6
21. Cauchois, C., Brassart, E., Drocourt, C., Vasseur, P.: Calibration of the omnidirectional vision sensor: Syclop. In: ICRA. (1999) 1287–1292
22. Kelley, C.T.: Solving Nonlinear Equations with Newton's Method. SIAM (2003) ISBN:0898715466.