# A Catadioptric Sensor with Multiple Viewpoints

Libor Spacek

*Department of Computer Science, University of Essex*
*Wivenhoe Park, Colchester, CO4 3SQ, UK*
*tel. +44 1206 872343, fax. +44 1206 872788*

**Abstract**

Conventional cameras with a limited field of view often lose sight of objects when their bearings change suddenly due to a significant turn of the observer (robot), the object, or both. Catadioptric omnidirectional sensors, consisting of a camera and a mirror, can track objects and estimate their distances more robustly.

The shapes of mirrors used by such sensors have differing merits. This paper discusses several advantages of the conical mirror over other shapes of mirrors in current use. A perspective projection unwarping of the conical mirror images is developed and demonstrated. This has hitherto been considered impossible for mirrors with multiple viewpoints.

An estimation of distance (range) over a large surrounding area is crucial in mobile robotics. A solution is proposed here in the form of an omnidirectional stereo apparatus with two catadioptric sensors in a vertical coaxial arrangement. The coaxial stereo requires very simple matching since the epipolar lines are the radial lines of identical orientations in both omnidirectional images. The radial matching is supported by a novel polar edge finder which uses discrete cosine transform and returns image gradients expressed in polar coordinates.

*Key words:* omnidirectional vision, catadioptric sensor, conical coaxial mirrors, stereopsis, radial edge finding

*Email address:* spacl@essex.ac.uk (Libor Spacek).
*URL:* cswww.essex.ac.uk/mv (Libor Spacek).

## 1 Introduction

This paper offers some reasoned suggestions for the best modality for autonomous mobile robots sensors. The introduction follows this sequence of arguments: amongst the senses, vision is usually the best choice; amongst vision sensors, omnidirectional sensors are the best for navigation and mapping; amongst omnidirectional vision sensors, catadioptric sensors are the best in dynamic environments; amongst catadioptric omnidirectional sensors, the conical mirror with a perspective camera offers the best image quality and resolution.

Passive visual sensors have known advantages in comparison with active sensors currently popular in much of robotics; specifically no mutual interference or detection and relatively accurate localisation of objects even at large distances. These are the overriding reasons why vision is the distal sense of choice for most biological organisms in air and clear water. If a robot is to emulate some aspects of (biological) mobility, then it needs, most of all, to emulate the (biological) visual abilities.

Some of the most important tasks in vision for robotics include autonomous navigation, site modelling (mapping) and surveillance. They all benefit from using panoramic 360° images produced by omnidirectional visual sensors.

Early attempts at using omnidirectional sensors included camera clusters (Swaminathan and Nayar, 2000) and various arrangements of mechanically rotating cameras and planar mirrors, (Rees, 1970), (Kang and Szeliski, 1997), (Ishiguro et al., 1992). These mostly had problems with registration, motion, or both. Fisheye lens cameras have also been used to increase the field of view (Shah and Aggarwal, 1997) but they proved difficult because of their irreversible distortion of nearby objects and the lack of a single viewpoint.

Single viewpoint projection geometry exists when the light rays arriving from all directions intersect at a single point known as the (single) effective viewpoint. For example, by placing the centre of the perspective camera lens at the outer focus of a hyperbolic mirror, the inner focus then becomes the single effective viewpoint.

Catadioptric sensors (Nayar, 1997) consist of a fixed dioptric camera, usually mounted vertically, plus a fixed rotationally symmetrical mirror suspended above or below the camera. The advantages of catadioptric sensors derive from the fact that, unlike the rotating cameras, their 'scanning' of the surroundings is almost instantaneous, the camera exposure time being usually shorter than the full circle mechanical rotation time. Shorter exposure time means fewer image capture problems caused by motion and vibration of the camera, or by moving objects.

Suitability for use in dynamic environments is clearly an important consideration, especially as one of the chief benefits of omnidirectional vision in general is the ability to retain objects in view even when their bearings have changed suddenly and significantly. Catadioptric omnidirectional sensors are therefore ideally suited to visual navigation (Rushant and Spacek, 1998), visual guidance applications (Pajdla and Hlavac, 1999), using stereopsis, motion analysis (Yagi et al., 1996), and site mapping (Yagi et al., 1995).

The main practical problem with catadioptric sensors is that the details of the image can have relatively poor resolution, as the image depicts a large area. The resolution problem is unfortunately compounded by mirrors whose shapes have curved cross-sections. Such radially curved mirrors include the three popular quadric surface mirrors (elliptic, hyperbolic and parabolic) which are known to possess a single viewpoint at their focal points.

A single viewpoint is generally thought to be necessary for an accurate unwarping of images and for an accurate perspective projection which is relied on by most current computer vision methods (Baker and Nayar, 1999). The single viewpoint projection has been endorsed and recommended by (Baker and Nayar, 1998, 2001), (Daniilidis and Geyer, 2000), (Geyer and Daniilidis, 2000a,b, 2002b), (Svoboda and Pajdla, 2002) and others.

There have been few attempts at analysing multi-viewpoint sensors (Swaminathan and Nayar, 2001), (Fiala and Basu, 2002), (Spacek, 2003), although various people (Yagi and Kawato, 1990) used them previously without analysis.

An omnidirectional sensor's resolution can be improved by using several planar mirrors with a separate camera for each one. The mirrors are placed in some spatial arrangement, for instance in a six sided pyramid (Yokoya et al., 1998). The mirrors are carefully adjusted so that all the reflected camera positions coincide and thus form a single effective viewpoint. However, such arrangements are awkward, expensive, and sensitive to alignment errors. The hexagonal pyramid apparatus would require no fewer than twelve precisely placed cameras for stereopsis! Also, the coverage of the surrounding area is not isotropic.

This paper proposes a solution to the above problems which combines the benefits of the planar mirrors (no radial distortion, no radial loss of resolution) with the advantages of the rotationally symmetric catadioptric sensor (short exposure, isotropic imaging). The only shape of mirror that satisfies these requirements is the cone.

Section 2 summarises the projection and the unwarping transformation for a single conical mirror.

Section 3 describes an omnidirectional stereo system using two coaxial conical mirrors and two cameras.

## 2 Perspective Projection through a Conical Mirror

The benefits of the cone mirror over the radially curved mirrors were pointed out by Lin and Bajcsy (2001). They can be summarised as:

(1) Curved cross-section mirrors produce inevitable radial distortions. Radial distortion is proportional to the radial curvature of the mirror. We note simply that the cone has constant zero radial curvature everywhere except at its tip point which will only be reflecting the camera anyway.
(2) Radially curved mirrors produce 'fish eye' effects: they magnify the objects reflected in the centre of the mirror, typically the camera, the robot, or the sky, all of which are of minimal interest. On the other hand, they shrink the region around the horizon, thereby reducing the available spatial resolution in the area which is of interest. See Figures 1 and 2 for the comparison of the hyperbolic and the conical mirrors. The mirrors are showing different scenes but both are pointing vertically upwards.
(3) The cone presents planar mirrors in cross-section. See Figure 3. The planar mirror does not distort the resolution density of the perspective camera.

Some optimised shapes of radially curved mirrors have been proposed (Hicks and Bajscy, 1999), as well as hybrid sensors, mirrors combining two shapes into one, and other mirrors of various functions. However, it seems that none of them completely address all of the above points.

The cone mirror has a single effective viewpoint located at the tip. However, this is a degenerate case causing practical difficulties with obtaining images. Lin and Bajcsy (2001) proposed cutting off the cone tip and placing the camera lens in its place, or placing the tip at the forward focus point of the lens. Both of these methods require the camera to be precisely positioned very close to the mirror which results in difficulties with capturing enough light and with focusing, so much so that the improvement in image quality over the radially curved mirrors is doubtful.

Our solution consists of placing the camera at an arbitrary comfortable distance $d$ and still obtaining a useful projection, despite the fact that there is now an infinite number of viewpoints arranged in a circle of radius $d$ around the tip of the cone. See Figure 4. Relaxing the precision of the camera distance represents an additional practical benefit in comparison with hyperbolic mirrors or the approach of Lin and Bajczy.

### *2.1 Field of View and Distance of View*

The maximum utilisable value $\phi_{max}$ of the camera's field of view angle is:

$$\phi_{max} = 2 * \arctan \frac{R}{R+d} \tag{1}$$

$R$ is both the radius and the height of the cone mirror with a 90° angle at the tip. Given the field of view angle $\phi$ of a particular camera lens, the appropriate camera distance $d$ is found by:

$$d = (cotan\frac{\phi}{2} - 1) \cdot R \tag{2}$$

Placing the camera at distance $d$ from the tip of the cone just encloses the base circle of the cone within the image. The field of view of the camera and the size of the mirror are thus utilised to their best advantage. For example, a mirror of radius $R = 60mm$ and a camera with $\phi = \pi/4$ results in $d = 85mm$ (rounded up). However, unlike for the focal mirrors, $d$ is not critical. At worst, one may lose a few pixels around the edges of the image.

### *2.2 The Projection*

Since the image of a rotationally symmetric mirror viewed along its axis of symmetry is circular, it is convenient to use the polar coordinates $(r_i, \theta)$ to represent the image positions and the related cylindrical coordinates $(r, \theta, h)$ for the 3D scene. See Figure 5 for the cross section in $\theta$ direction of the perspective projection via the conical mirror. Suppose we are projecting scene point P located at the 3D coordinates $(r, \theta, h)$. Note that the four points $P$, $C$, the centre of the lens, and the image projection of $P$ all lie on the same projection ray and are therefore collinear (forming similar triangles).

Let us denote the image radius value of the projection of $P$ by $h_i$ (the image height of $P$). The perspective projection formula needs to relate $h_i$ to $h$. It is obtained from the collinearity property (or two similar triangles in Figure (5)):

$$h_i = \frac{v \cdot h}{d + r} \tag{3}$$

$h_i$ values are always positive (image radius). This is equivalent to using front projection to remove the image reversal. Equation (3) is much simpler than the corresponding projection equations for the radially curved mirrors.

$v$ is the distance of the image plane behind the centre of the thin lens in Gaussian optics. This model is also valid for thick lenses, which have a cylindrical section inserted between two spherical surfaces. In that case $d$ would denote the distance to the front principal point of the lens (start of the cylindrical section) and $v$ would be measured from the back principal point (end of the cylindrical section). The focal length is normally less than $v$, unless we reduce $v$ to focus on infinity, or use the simplifying pinhole camera assumption. The calibration of $v$ is obtained by substituting $r_m$ for $h_i$ and $R$ for both $h$ and $r$ in equation (3):

$$v = (\frac{d}{R} + 1) \cdot r_m \tag{4}$$

The image radius of the mirror $r_m$ is determined by locating the outer contour of the mirror in the image by using Hough transform or other methods. Suppose it is found that $r_m = 100$ pixel units. Substituting the values for $d$ and $R$ from our previous example, we get $v = 242$ pixel units (rounded up). $v$ in (3) now acts as the conversion constant between $h$ in $mm$ units and $h_i$ in pixel units.

The perspective projection function using the single effective viewpoint at the tip of the mirror at $(0, 0, 0)$ is just a special case of equation (3), where $d = 0$. Suppose that a thought-experiment (Gedanken) world exists in which all the objects are pushed distance $d$ further away from the mirror axis. The single viewpoint projection of such world would result in the same image as our multiple viewpoint projection of the real world.

It is also clear that once $r$ is known (see the stereopsis method below), it is possible to reconstruct the single viewpoint projection of the real world by using equation (3) and setting $d = 0$.

The relationship between these two projection geometries is illustrated in Figure 4, using the concept of virtual projection cylinders with the same axis as the cone. The projection cylinder for the single viewpoint at $(0, 0, 0)$ has the radius $v$ (the innermost circle in Figure 4). The projection cylinder for the multiple viewpoints, depicted by the outermost circle in the top view cross-section in Figure 4, has the radius $d + v$. This is the only difference between the two projections.

So far, we considered the projection for a fixed value of $\theta$ and identified its associated viewpoint. Now we fix the elevation angle $\epsilon = \arctan(h/(d + r))$ and allow $\theta$ to vary.

Imagine spinning Figure 5 around the mirror axis. All projection lines with the same elevation angle will intersect the cone axis at the single point $C$ at

coordinates $(0, \theta, h_c)$. Thus the intersection point $C$ is the viewpoint associated with the elevation $\epsilon$.

We can determine $h_c$ of $C$ from $h$ of $P$ by again using the collinearity property:

$$h_c = \frac{(d \cdot h)}{(d + r)} \tag{5}$$

Sensors with a single (global) effective viewpoint have perspective projections of the same form in both orthogonal image dimensions (usually $x, y$). However, we get a different perspective projection in the $\theta$ dimension, as the effective viewpoint $C$ for the $\theta$ dimension is different from the effective viewpoint $(d, \theta + \pi, 0)$ for the $h$ dimension, which is the centre of the lens of the reflected camera as discussed earlier.

Specifically, suppose that the size of some 3D object in $\theta$ dimension is $w$ (ie. the object's width). The image width $w_i$, as projected onto the virtual projection cylinder, is given by the following perspective projection formula:

$$w_i = \frac{(v + d) \cdot w}{r} \tag{6}$$

Equation (6) is not needed for our stereopsis which uses only the $h$ projection but it could be utilised if we placed two mirrors side-by-side. It has been used in this fashion by Brassart et al. (2000).

We now define the projection property whereby the viewpoints are said to be dimensionally separable:

- Each radial line in the image (or equivalently each column in the unwarped image) has its own unique viewpoint.
- Each concentric circle in the image (or equivalently each row in the unwarped image) has its own unique viewpoint.
- Each pixel is aligned with its two (row and column) viewpoints, along the projection line from $P$.

### *2.3 Registration*

We have just described the idealised projection which will be valid and accurate after registration, when the tip of the mirror is precisely aligned with the centre of the image and the axis of view coincides with the axis of the mirror. In general, registration needs to be performed to find the two translation and three rotation parameters needed to guarantee this. Existing registration methods will also apply and work in this situation. Geyer and Daniilidis

(2001, 2002a) present good solutions to this problem within the context of omnidirectional vision.

Straight lines in the 3D world become generally conic section curves when projected. However, lines which are coplanar with the axis of the mirror will project into radial lines. Concentric circles around the mirror will project again into concentric circles. These properties can be utilised for a simple test card registration method, where the test card is of the 'shooting target' type consisting of cross-hairs and concentric circles, centered on the cone axis.

### *2.4 Unwarping of the Input Image*

Cutting and unrolling the virtual projection cylinder results in the required unwarped rectangular panoramic image $g(x, y)$. Therefore unwarping is the backprojection of the input image $f(x_i, y_i)$ onto the virtual projection cylinder. The mapping from coordinates of $g(x, y)$ to the polar coordinates $(h_i, \theta_i)$ of the input image is:

$$h_i = \frac{r_m}{y_m} y \quad , \quad \theta_i = \frac{2\pi}{x_m} x \tag{7}$$

where $(x_m, y_m)$ are the desired dimensions of $g(x, y)$ in pixel units, $r_m$ is the radius of the mirror as seen in the input image in pixel units, and $\theta_i$ is in radians. The correct aspect ratio of $g(x, y)$ is: $x_m/y_m = 2\pi$, based on the perimeter of the mirror which maps to the bottom row of the unwarped image.

In order to generate the correct pixel values of $g(x, y)$, we need the mapping (lookup) from any given $(x, y)$ to the corresponding sub-pixel position $(x_i, y_i)$ in the original rectangular coordinates of the input image. We use polar coordinates as an intermediate step, and then equations (7). We also need to find the centre of the mirror in the input image $(x_c, y_c)$.

$$x_i = x_c + h_i \cos\theta_i = x_c + \frac{r_m}{y_m} y \cdot cos(\frac{2\pi}{x_m} x) \tag{8}$$

$$y_i = y_c + h_i \sin\theta_i = y_c + \frac{r_m}{y_m} y \cdot sin(\frac{2\pi}{x_m} x) \tag{9}$$

This can be used as it stands by simply assigning the nearest input image pixel $f(round(x_i), round(y_i))$ to $g(x, y)$.

We implemented the unwarping using two dimensional DCT (discrete cosine transform) of the omnidirectional input image, instead of the usual less precise

pixel rounding or interpolation. The main advantage of this approach becomes apparent when performing the polar edge-finding in the next section.

We use the forward DCT as follows:

$$c(q,p) = \frac{a(q,p)}{XY} \sum_{y_i=0}^{Y-1} \sum_{x_i=0}^{X-1} f(x_i, y_i) \cdot cos(\frac{\pi q}{Y}(y_i + 0.5)) \cdot cos(\frac{\pi p}{X}(x_i + 0.5)) \quad (10)$$

where: $a(q,p) = 1$ when $q = p = 0$; $a(q,p) = 2$ when $q \neq p$ and $qp = 0$; $a(q,p) = 4$ otherwise. This definition of $a(q,p)$ allows us to leave it out of the inverse DCT. $X, Y$ are the dimensions of the discrete input image $f(x_i, y_i)$. $c(q,p)$ is the normalised coefficients array of dimensions $P, Q$ produced by the forward DCT.

Inverse DCT:

$$f(x_i, y_i) = \sum_{q=0}^{Q-1} \sum_{p=0}^{P-1} c(q,p) \cdot cos(\frac{\pi q}{Y}(y_i + 0.5)) \cdot cos(\frac{\pi p}{X}(x_i + 0.5)) \quad (11)$$

Choosing $P = \frac{X}{4}$ and $Q = \frac{Y}{4}$ for the coefficients array dimensions usually gives nearly perfect continuous fit to the discrete input data.

Using equations (8) and (9) to substitute for $x_i$ and $y_i$ in the right hand side of equation (11), we get the unwarping function:

$$g(x,y) = f(x_i, y_i) = \sum_{q=0}^{Q-1} \sum_{p=0}^{P-1} c(q,p) cos(\frac{\pi q}{Y}(y_c + \frac{r_m}{y_m} y \cdot sin(\frac{2\pi}{x_m} x) + 0.5)) \cdot$$
$$cos(\frac{\pi p}{X}(x_c + \frac{r_m}{y_m} y \cdot cos(\frac{2\pi}{x_m} x) + 0.5)) \quad (12)$$

This evaluates the now continuous image function $f(x_i, y_i)$ exactly at the required (sub-pixel) position. To sum up, the DCT unwarping consists of the following steps:

(1) Perform the forward DCT transform on the input image, using equation (10) to compute $c(q,p)$.
(2) Generate the unwarped rectangular image $g(x,y)$ by evaluating equation (12) in the usual $x, y$ scanning order.
(3) For colour images, this process should be applied separately to each colour plane $(r, g, b)$.

*2.5 Unwarping Results and Discussion*

Figure 6 shows the unwarping applied to the hyperbolic mirror image in Figure 1. Figure 7 shows the unwarping of the conical mirror image in Figure 2.

Note that the conical mirror image utilises the available vertical resolution of the image better. About half way up the image in Figure 7 is where most of the extra resolution is gained in comparison with Figure 6. This provides better resolution of nearby objects for visual robot guidance and for stereopsis.

Note from the paved area that the correct perspective view of the curves and lines has been restored.

The $\theta$ resolution near the tip of the mirror (top of the unwarped image) is clearly limited: any registration errors will produce distortions at the top of the unwarped image, as shown in Figure 7. Even if a few pixel rows at top of the image become quite useless, this represents fewer wasted pixels than the typical image area of the sky in the hyperbolic mirror images.

It is perhaps better to point the tip of the mirror downwards as in our diagrams, so that the available image resolution increases towards the (higher) horizon. The ratios of the number of available pixels per unit of length on the depicted objects at various distances are then better balanced.

## 3 Coaxial Omnidirectional Stereo

Various arrangements have been proposed for binocular systems using catadioptric sensors. Two mirrors situated side by side can be used to compute the distance of objects in terms of the disparity measured as the arising difference in angles $\theta$ (Brassart et al., 2000). However, such arrangement is not truly omnidirectional, as a large part of the scene will be obstructed by the other catadioptric sensor.

It is better to arrange the cameras coaxially to avoid this problem. The coaxial arrangement has the further major advantage of having simple aligned radial epipolar lines. Lin and Bajcsy (2003) used a single conical mirror and attempted to place two cameras at different distances along its axis. They had to use a beam-splitter to avoid the nearer camera obstructing the view of the more distant camera. See Figure 8.

We propose an omnidirectional stereo system consisting of two coaxial conical mirrors pointing in the same direction, each with its own camera, as outlined in Figure 9.

### *3.1 Radial Triangulation*

We wish to obtain a triangulation formula for the radial distance of objects. The radial distance $r$ is measured from the axis of the mirror(s) to any 3D scene point $P$, which has to be in the region that is visible by both cameras (the common region). See Figure 9. The common region is annular in shape in 3D, with a triangular cross-section extending to infinity. It is bounded above and below in the $(r, h)$ plane by the lines: $h = \frac{(r+d)\cdot R}{d+R}$, and $h = s$. The angle at the tip of the common region triangle is $\frac{\phi}{2}$. Let the distance of the nearest point (the tip of the common region) be $r_{min}$. Stereopsis cannot be employed anywhere nearer than $r_{min}$:

$$r_{min} = s \cdot (\frac{d}{R} + 1) - d \tag{13}$$

In order to obtain the triangulation formula, we use two instances of equation (3) for two coaxial mirrors separated by distance $s$ between them ($s$ is measured along the h axis). We assume here that the parameters $v$ and $d$ are the same for both cameras, though this can be easily generalised if necessary.

$$(d + r)h_{i1} = v(h - s) \tag{14}$$

$$(d + r)h_{i2} = vh \tag{15}$$

Subtracting (14) from (15) and manipulating a little, we obtain the triangulation formula:

$$r = \frac{vs}{h_{i2} - h_{i1}} - d \tag{16}$$

This is very similar to the usual triangulation formula from classical side-by-side stereopsis but here the disparity $(h_{i2} - h_{i1})$ is radial. Note that the extra distance d of the reflected camera is correctly subtracted out. The familiarity of our formula is not surprising, as the two reflected cameras form a classical stereo system which happens to have a vertical baseline.

### *3.2 Polar Edge Finding and Radial Matching*

The edges are located by our polar edge-finder using the inverse DCT and the polar coordinates $(h_i, \theta_i)$ of the input image. This is a convenient way to compute the partial derivatives of the input image in $h_i$ and $\theta_i$ directions.

The polar edge finding approach entirely avoids the slow unwarping process. The unwarping is needed only for the convenience of human viewing, such as in Figure 10, showing a traditional edge map of the unwarped image, using (Spacek, 1986). In this example, a naive coaxial stereopsis implementation would match edges with close to vertical image gradients along the vertical epipolar lines. Our faster direct method matches the (same) edges with nearly radial gradients along the radial epipolar lines in the original (warped) input images.

We now develop the first derivative polar edge finder. Similar process can be followed to apply higher derivatives or other functions to the input image. Substituting the polar coordinates from equations (8),(9) for $x_i, y_i$ in the right hand side of equation (11), we get the inverse DCT in polar form:

$$f(x_i, y_i) = \sum_{q=0}^{Q-1} \sum_{p=0}^{P-1} c(q,p) cos(\frac{\pi q}{Y}(y_c + h_i sin(\theta_i) + 0.5)) \cdot \\ cos(\frac{\pi p}{X}(x_c + h_i cos(\theta_i) + 0.5)) \tag{17}$$

We can differentiate equations (11) or (17) instead of differentiating the input image. This is legitimate as the inverse DCT has finite number of terms $PQ$. Differentiating with respect to $h_i$ we get:

$$\frac{\partial f(x_i, y_i)}{\partial h_i} = -\sum_{q=0}^{Q-1} \sum_{p=0}^{P-1} \frac{\pi \cdot c(q,p)}{\sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}} \cdot \\ \{\frac{q(y_i - y_c)}{Y} \cdot sin(\frac{\pi q}{Y}(y_i + 0.5)) \cdot cos(\frac{\pi p}{X}(x_i + 0.5)) + \\ \frac{p(x_i - x_c)}{X} \cdot sin(\frac{\pi p}{X}(x_i + 0.5)) \cdot cos(\frac{\pi q}{Y}(y_i + 0.5))\} \tag{18}$$

Differentiating with respect to $\theta_i$ produces the second component of the polar image gradient:

$$\frac{\partial f(x_i, y_i)}{\partial \theta_i} = -\sum_{q=0}^{Q-1} \sum_{p=0}^{P-1} \pi \cdot c(q,p) \cdot \\ \{\frac{q(x_i - x_c)}{Y} \cdot sin(\frac{\pi q}{Y}(y_i + 0.5)) \cdot cos(\frac{\pi p}{X}(x_i + 0.5)) - \\ \frac{p(y_i - y_c)}{X} \cdot sin(\frac{\pi p}{X}(x_i + 0.5)) \cdot cos(\frac{\pi q}{Y}(y_i + 0.5))\} \tag{19}$$

The polar edge finding consists of the following steps:

(1) Perform forward DCT transform of the input image, using equation (10).
(2) Use equations (18),(19) to find the polar image gradient vector at any image position.

We now have a global continuous gradient function (a polar edge map) of the input image. It follows that it is not necessary to generate edge maps of the whole images when doing the stereo matching. The image gradient can be evaluated on demand at any sub-pixel point. The outline of the radial stereo matching algorithm is as follows:

(1) Given a pair of stereo images $f_1$ and $f_2$, find all significant points in $f_1$ where $abs(\frac{\partial f_1}{\partial h_i})$ passes some threshold ($abs()$ is the absolute value function). Store the $\frac{\partial f_1}{\partial h_i}$ values at such points.
(2) At each significant point, evaluate and store $\frac{\partial f_1}{\partial \theta_i}$ as well.
(3) Select the next significant point $\mathbf{s}$ in $f_1$ and note its $\theta_i$ value.
(4) Find and store, in the same way as above, all significant points along the radial line of $\theta_i$ orientation in $f_2$. If done before for this $\theta_i$, retrieve from memory instead.
(5) Find the best match for $\mathbf{s}$ along this line, looking for the most similar image gradient vector $(\frac{\partial f_2(h_i,\theta_i)}{\partial h_i}, \frac{\partial f_2(h_i,\theta_i)}{\partial \theta_i})$. Pay attention to sensible ordering of the matches plus any other stereopsis matching constraints, e.g. the interest horizon of the next section.
(6) Compute the distance of object for the successful match, using the obtained values $h_{i1}$ and $h_{i2}$ and the triangulation equation (16).
(7) Repeat from 3.

There are other sophisticated stereo matching methods that could be adapted to these circumstances, for example (Sara, 2002).

### *3.3 Stereopsis Error Analysis*

Image position errors are produced by image discretisation (pixelation) and feature finding (edge localisation). Denote the resulting imaged position error as a displacement $dh_i$ to $h_i$. Given $dh_i$, what is the stereopsis distance error $dr$?

Let $p = h_{i2} - h_{i1}$ be the disparity; then $dp = dh_{i2} - dh_{i1}$. Substitute $p$ for the denominator in equation (16), differentiate w.r.t. $p$, and simplify using $\frac{vs}{p} = d + r$:

$$dr = -\frac{vs}{p^2}dp = -\frac{d+r}{p}dp = (d+r)\frac{dh_{i1} - dh_{i2}}{h_{i2} - h_{i1}} \tag{20}$$

Sometimes the stereo matcher selects a wrong target, which results in a significant and essentially unpredictable error $dp$. The occurances of this error are reduced primarily by ensuring the stability and uniqueness of the image features descriptions, which are common aims of most good vision systems.

The $p^2$ term in the denominator of equation (20) suggests eliminating stereo matches whereby $p^2$ falls below some threshold. This is equivalent to imposing an *interest horizon*, whereby all points that are too distant are eliminated even before their distance was explicitly found!

We tested the stereopsis errors on artificially generated 3D scenes of random shapes and points with known ground truth $r_{true}$. Omnidirectional 2D images were then generated by a graphics projection of those scenes using the perspective equation (3). Finally, the created images were used to compute $r_{stereo}$ as described in this paper ($dr = r_{stereo} - r_{true}$). Admittedly, such testing is not as demanding as using real images to start with, particularly with regard to the stereopsis matching.

### *3.4 Stereopsis Discussion*

Each sensor consists of one conical mirror and one ordinary perspective camera. Each sensor's visible region is bounded by the plane perpendicular to the axis of the mirror and touching the tip of the mirror. The visible region is on the same side of the plane as the mirror. The image resolution increases with the elevation angle $\epsilon$.

The illustrated stereo arrangement (both cone tips pointing down) projects the best image resolution at the edge(s) of the mirror(s) at the elevated objects on the horizon. The horizon is normally rich in natural visual features of high contrast that make it useful for general outdoors navigation (Rushant and Spacek, 1998). For visual guidance indoors, the entire stereo apparatus can be simply inverted (turned upside down). The best resolution will then be directed at the nearest objects on the ground, which is more useful for accurate stereopsis. Very close range stereo with narrow common region centered on the mid-point between the mirrors is obtained by making the tips of the mirrors face away from each other. In each case the triangulation and matching will be the same. The only combination to be avoided for stereopsis is where the tips of the mirrors are facing each other, as this results in no common region being visible by both cameras.

Our arrangement is quite different from that proposed by Lin and Bajczy. The resulting triangulation formula is different. Our system is simpler, there is no loss of light through the beam splitter, and we gain better image quality by being able to view large size conical mirrors.

It is possible to generalise the projection equations to a cone with any angle at the tip. The dimensional separability property of the viewpoints still holds. The virtual projection cylinder becomes a cone. There are image quality (focusing) advantages to using flatter cone mirrors with $\alpha > 90°$. This is especially to be recommended for smaller scenes, such as those typically used in obstacle avoidance.

The general benefits of the coaxial omnidirectional stereopsis are both practical (objects do not disappear from view due to vehicle rotation), and theoretical/computational (the epipolar geometry is simpler than in classical stereopsis).

The presented polar edge finder methodology should be of interest to omnidirectional vision generally, as it can be used with any rotationally symmetric mirrors. We demonstrated that the unwarping is made unnecessary by working in polar coordinates.

## 4 Conclusion

This paper has identified the conical mirror as a good solution for catadioptric omnidirectional sensors.

The benefits of conical mirrors had been hitherto mostly overlooked because of the demand for a single viewpoint projection. This has resulted in the general use of the hyperbolic mirror viewed by a perspective camera from a precise focal distance, or of other quadric focal mirrors of similar shapes.

Three theoretical arguments were put forward here in support of our multi viewpoint projection. First, the multi viewpoint projection was shown to be equivalent to the single viewpoint projection of a more distant scene. Second, once stereopsis has been solved, it is possible to project the points of interest to a new image through a virtual single viewpoint. Third was the new concept of dimensionally separable viewpoints.

We conclude that the single viewpoint is not necessary for a useful perspective projection when using the conical mirror as described in this paper.

This means that such conical mirrors of any size provide a useful model of projection when viewed from any reasonable distance by any ordinary perspective camera. Conical mirrors are less sensitive to the precise distance of the camera than are hyperbolic, elliptic, or other focal mirrors. The ability to view the mirror from a greater distance allows the use of larger mirrors with relatively better optical quality. Given the same physical surface quality

(roughness), the optical quality will be proportional to the dimensions of the mirror.

The radial distortion properties of conical mirrors are better when compared to other circular mirrors.

Last but not least, conical mirrors direct the camera resolution into more useful parts of the surroundings and their resolution density is well behaved. In particular, conical mirrors deflect the central (foveal) projection beams towards the horizon. Hyperbolic (and other 'blunt tip') mirrors reflect the foveal beams back towards the camera. Therefore the conical mirror makes better use of the available image pixels.

The unwarping methods and experiments demonstrated the concept of an accurate perspective projection via multiple viewpoints and provided a solid basis for the stereopsis computation.

Apart from the coaxial omnidirectional stereopsis system presented here, this sensor is very suitable for a number of other interesting activities which are currently under investigation, such as: virtual reality, active vision, and omnidirectional optic flow for autonomous vehicle and robot guidance.

## References


Baker, S., Nayar, S., 1998. A theory of catadioptric image formation. In: ICCV98. pp. 35–42.

Baker, S., Nayar, S., November 1999. A theory of single-viewpoint catadioptric image formation. IJCV 32 (2), 175–196.

Baker, S., Nayar, S., 2001. Single viewpoint catadioptric cameras. In: PV01. pp. 39–71.

Brassart, E., et al., June 2000. Experimental results got with the omnidirectional vision sensor: Syclop. In: EEE Workshop on Omnidirectional Vision (OMNIVIS'00). pp. 145–152.

Daniilidis, K., Geyer, C., 2000. Omnidirectional vision: Theory and algorithms. In: ICPR00. Vol. 1. pp. 89–96.

Fiala, M., Basu, A., 2002. Panoramic stereo reconstruction using non-svp optics. In: ICPR02. Vol. 4. pp. 27–30.

Geyer, C., Daniilidis, K., 2000a. Equivalence of catadioptric projections and mappings of the sphere. In: OMNIVIS00. pp. xx–yy.

Geyer, C., Daniilidis, K., 2000b. A unifying theory for central panoramic systems and practical applications. In: ECCV00. pp. xx–yy.

Geyer, C., Daniilidis, K., 2001. Structure and motion from uncalibrated catadioptric views. In: CVPR01. Vol. 1. pp. 279–286.

Geyer, C., Daniilidis, K., April 2002a. Paracatadioptric camera calibration. IEEE PAMI 24 (4), 1–10.

Geyer, C., Daniilidis, K., 2002b. Properties of the catadioptric fundamental matrix. In: ECCV02. Vol. 2. p. 140 ff.

Hicks, A., Bajscy, R., June 1999. Reactive surfaces as computational sensors. In: The second IEEE Workshop on Perception for Mobile Agents. Held in Conjunction with CVPR'99. pp. 82–86.

Ishiguro, H., Yamamoto, M., Tsuji, S., February 1992. Omni-directional stereo. PAMI 14 (2), 257–262.

Kang, S., Szeliski, R., November 1997. 3-d scene data recovery using omnidirectional multibaseline stereo. IJCV 25 (2), 167–183.

Lin, S., Bajcsy, R., 2001. True single view point cone mirror omni-directional catadioptric system. In: ICCV01. Vol. 2. pp. 102–107.

Lin, S., Bajcsy, R., May 2003. High resolution catadioptric omni-directional stereo sensor for robot vision. In: IEEE International Conference on Robotics and Automation, Taipei, Taiwan. pp. 12–17.

Nayar, S., 1997. Catadioptric omnidirectional cameras. In: CVPR97. pp. 482–488.

Pajdla, T., Hlavac, V., 1999. Zero phase representation of panoramic images for image vased localization. In: Computer Analysis of Images and Patterns. pp. 550–557.
URL citeseer.nj.nec.com/pajdla99zero.html

Rees, D., April 1970. Panoramic television viewing system. US Patent No. 3,505,465.

Rushant, K., Spacek, L., January 1998. An autonomous vehicle navigation system using panoramic vision techniques. In: International Symposium on Intelligent Robotic Systems, ISIRS98. pp. 275–282.

Sara, R., 2002. Finding the largest unambiguous component of stereo matching. In: ECCV (3). pp. 900–914.
URL citeseer.nj.nec.com/sara02finding.html

Shah, S., Aggarwal, J., 1997. Mobile robot navigation and scene modeling using stereo fish-eye lens system. MVA 10 (4), 159–173.

Spacek, L., February 1986. Edge detection and motion detection. Image and Vision Computing 4 (1), 43–56.

Spacek, L., June 2003. Omnidirectional perspective and stereopsis with conical mirrors. Research Report CTU–CMP–2003–12, Center for Machine Perception, K13133 FEE Czech Technical University, Prague, Czech Republic.

Svoboda, T., Pajdla, T., August 2002. Epipolar geometry for central catadioptric cameras. IJCV 49 (1), 23–37.

Swaminathan, R. Grossberg, M., Nayar, S., 2001. Caustics of catadioptric cameras. In: ICCV02.

Swaminathan, R., Nayar, S. K., 2000. Nonmetric calibration of wide-angle lenses and polycameras. IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (10), 1172–1178.
URL citeseer.nj.nec.com/swaminathan00nonmetric.html

Yagi, Y., Kawato, S., 1990. Panoramic scene analysis with conic projection. In: IROS90.

Yagi, Y., Nishii, W., Yamazawa, K., Yachida, M., 1996. Rolling motion estimation for mobile robot by using omnidirectional image sensor hyperomnivision. In: ICPR96.

Yagi, Y., Nishizawa, Y., Yachida, M., October 1995. Map-based navigation for a mobile robot with omnidirectional image sensor copis. Trans. Robotics and Automation 11 (5), 634–648.

Yokoya, N., Iwasa, H., Yamazawa, K., Kawanishi, T., Takemura, H., 1998. Generation of high-resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors. In: ICPR98.

Fig. 1. *An omnidirectional image obtained using a hyperbolic mirror and an ordinary perspective camera. Note the typical predominance of the sky in the image.*

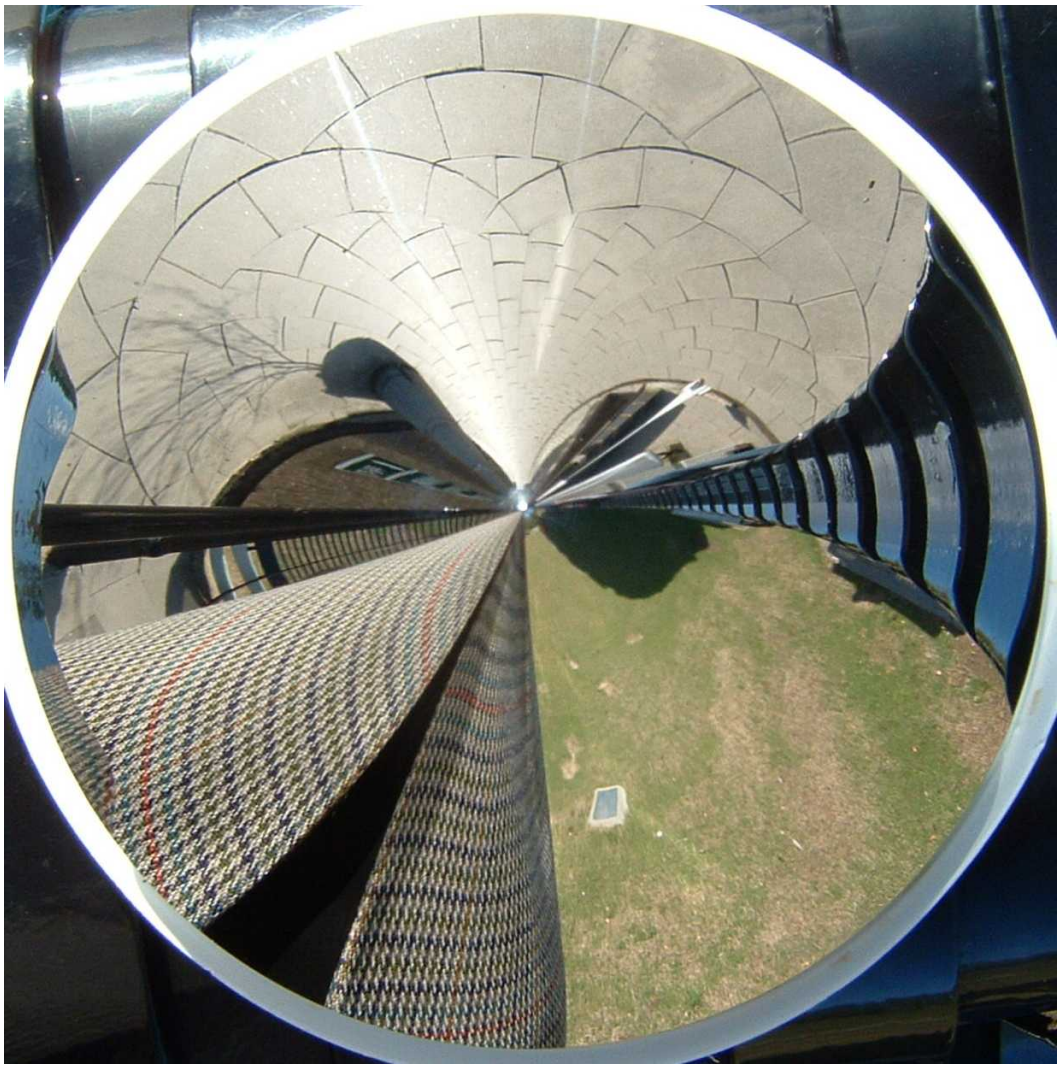

Fig. 2. *A conical mirror image showing a grass area, a paved area, and a part of a jacket, taken with an ordinary perspective camera. The entire mirror image depicts useful data. The tip of the cone pointed upwards.*

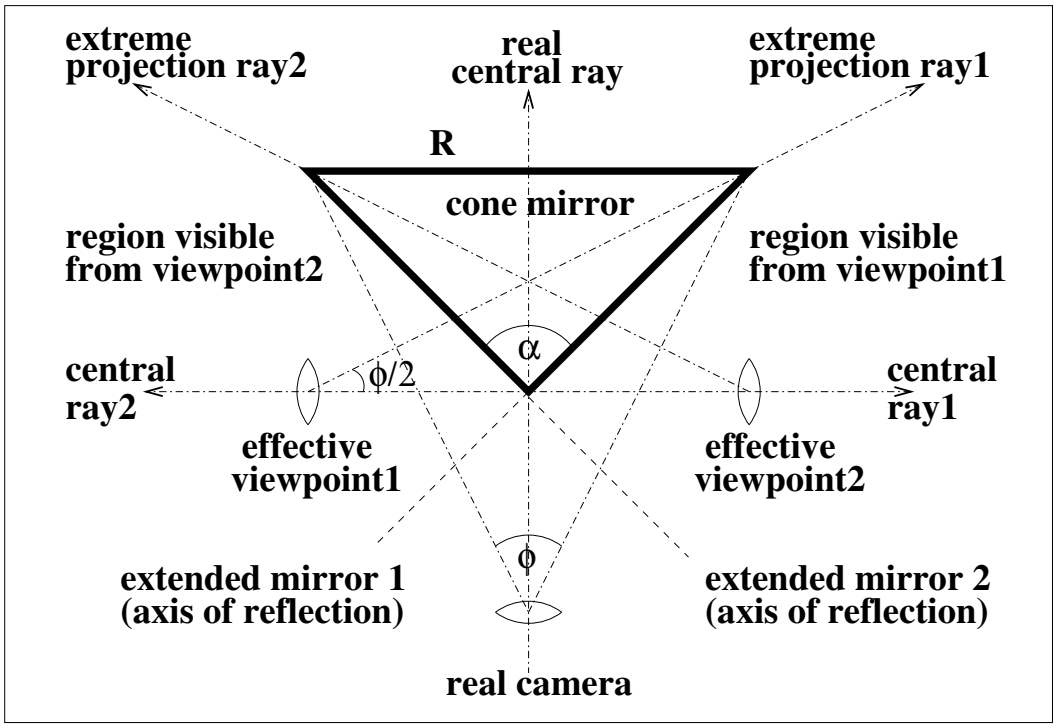


Fig. 3. *Cross section of the conical mirror projection geometry. According to the laws of optics, mirrors can reflect either the objects or the viewpoints. The two situations are equivalent. In this case, the real camera with a field of view $\phi$ is reflected in two planar mirrors, creating two effective viewpoints. Each viewpoint has a field of view $\phi/2$ between its central projection ray and its extreme ray. The angle at the tip of the cone is $\alpha = 90°$ to ensure that the two effective lines of sight (central rays 1&2) are oriented directly towards each other. R is the radius of the mirror.*

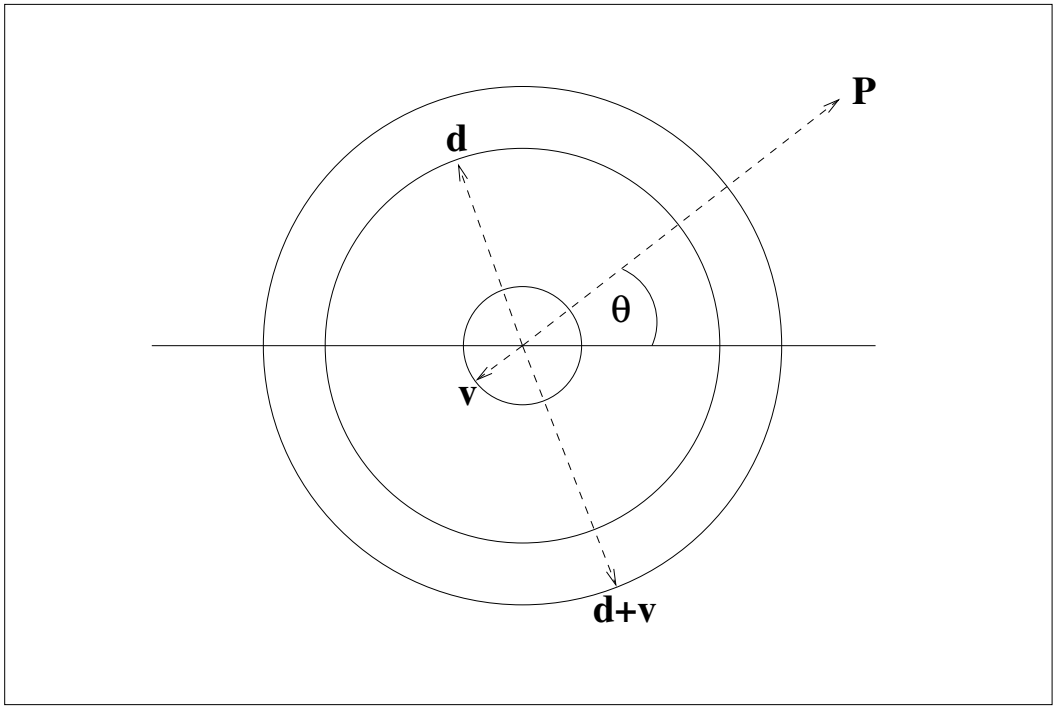


Fig. 4. *Schematic top view of the perspective projection of P via the conical mirror. The inner circle of radius v represents the virtual projection cylinder associated with the single viewpoint at the tip of the mirror. The circle of radius d is the locus of the viewpoints of the real camera at distance d. The outermost circle of radius d+v is the projection cylinder associated with the real camera.*

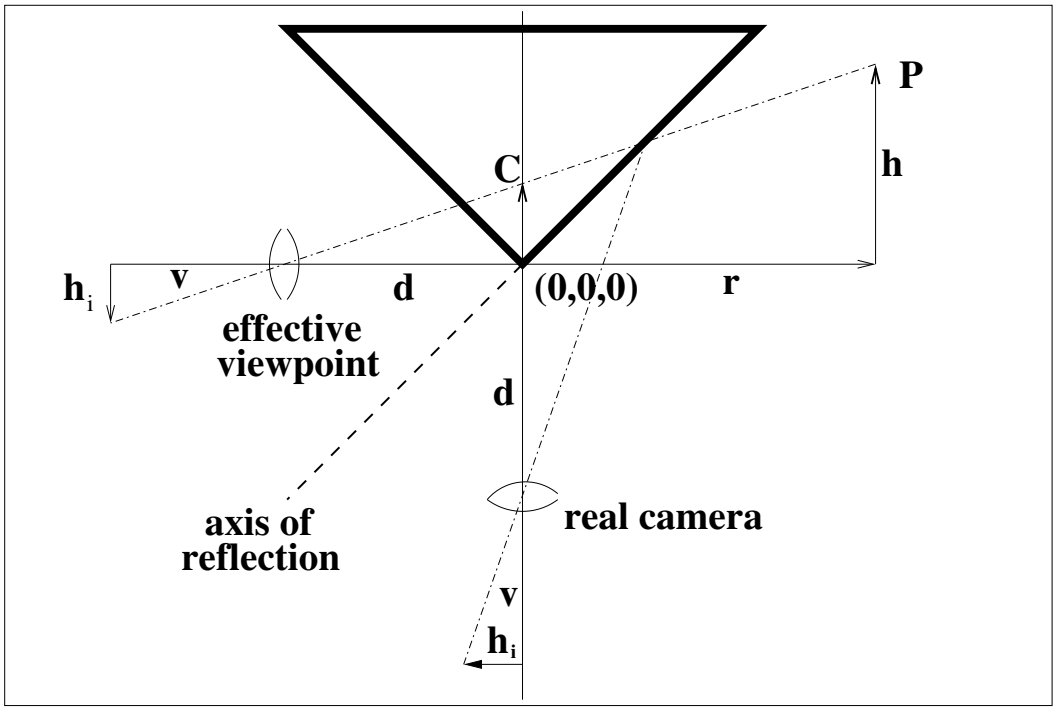


Fig. 5. *Cross section of the perspective projection of P via the conical mirror: d is the distance from the tip of the cone to the centre of the thin lens, v is the distance from the centre of the thin lens to the image.*

Fig. 6. *Unwarping of Figure 1.*

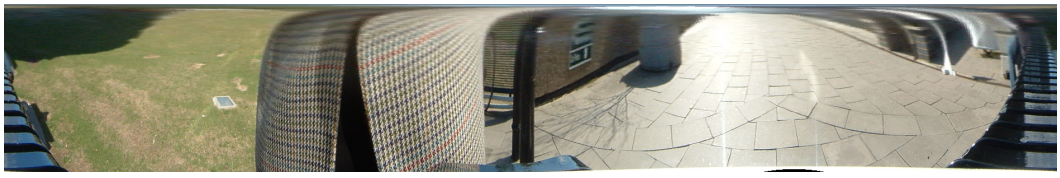

Fig. 7. *Unwarping of Figure 2 showing the correct linear perspective of the straight edges on the paving.*

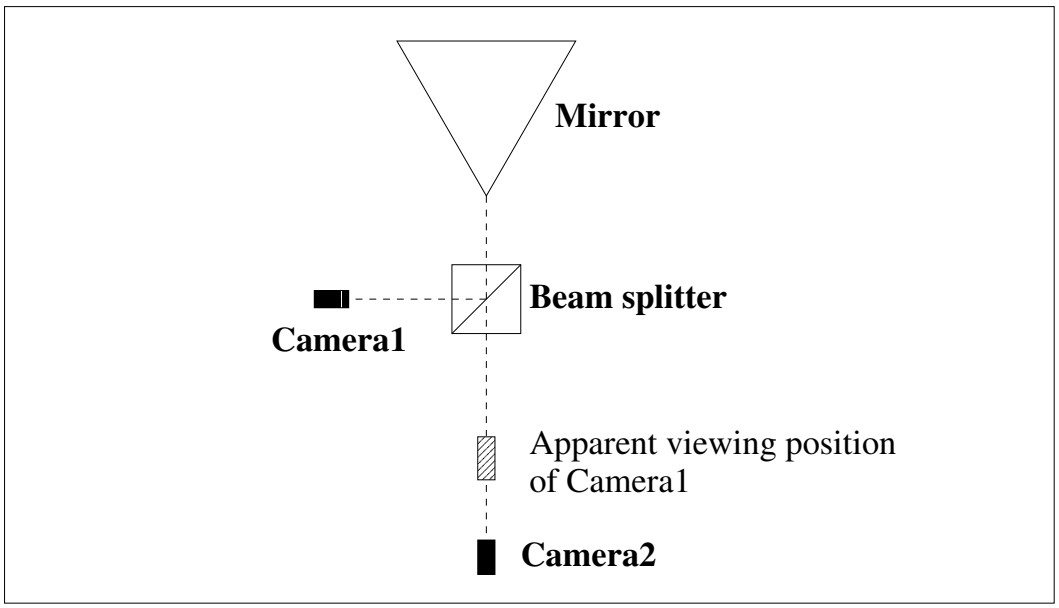


Fig. 8. *Lin and Bajcsy's omnidirectional stereo using a single conical mirror and two cameras at different distances. The beam splitter avoids an obstruction of the second camera's view but reduces the amount of available light.*

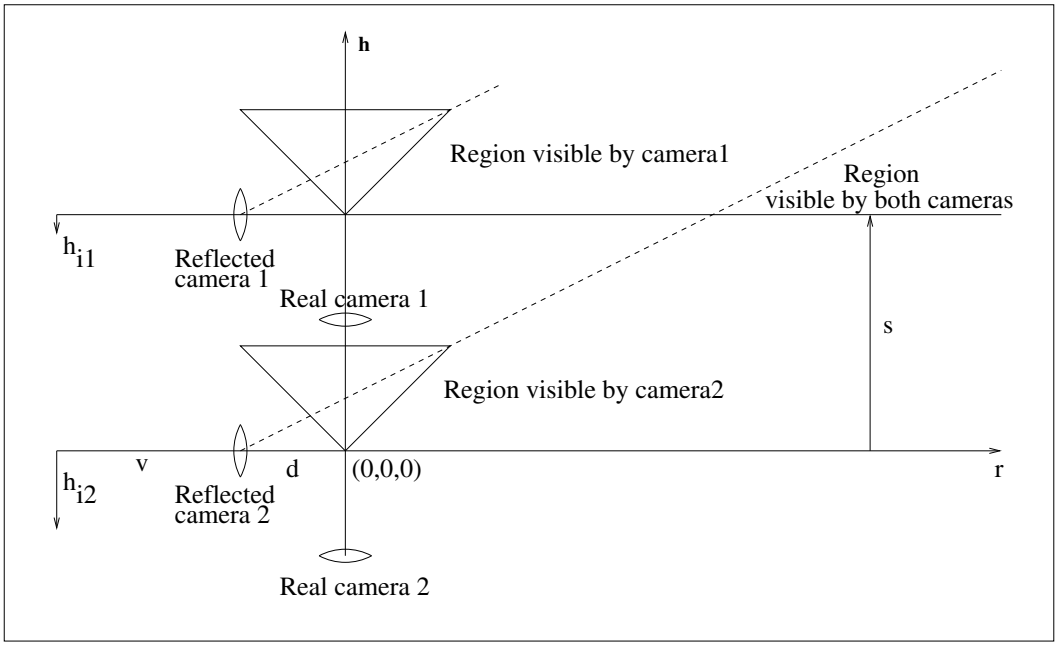


Fig. 9. *Omnidirectional stereo using two coaxial mirrors.*

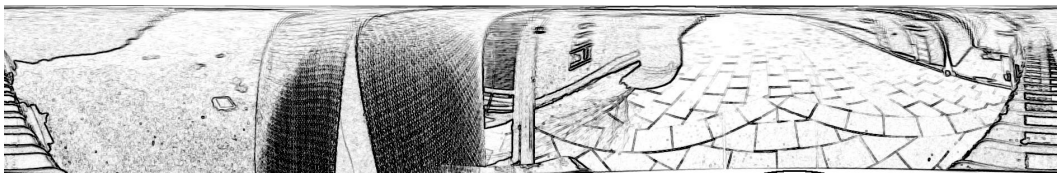

Fig. 10. *Edge map of the unwarped image in Figure 7.*