Range Recovery using Virtual Multi-camera Stereo

David W. Murray and Paul A. Beardsley

Robotics Research Group,
Department of Engineering Science,
Parks Road, Oxford University, OX1 3PJ, U.K.

Abstract

We describe the principles of a device comprised of a static camera and rotating plane mirror that enables the passive recovery of 3D range information. The range recovery can be regarded as either structure from known motion or as virtual multi-camera stereo. Two advantages of the arrangement are that the camera-mirror system is compact and that range information can be recovered over a wide, near panoramic, field of view.

1 Introduction

The most powerful methods of depth recovery using passive vision, shape from stereo and shape from motion, both involve obtaining images of a scene from different viewpoints, where the change of viewpoint must involve a translation of the optic centre of the camera. At the level of mechanism, this requirement detracts somewhat from the elegance of vision as a simple sensor. Either one has to have two (or more) cameras able to provide the different viewpoints simultaneously, or one has to have a device to move the camera to its new viewpoint. Both routes result in relatively complex and bulky apparatus.

We were concerned to find whether there were ways of making viewpoint displacements more simply to produce a compact, mechanically neat, sensor. The solution we present here exploits the near perfect specular reflection of broadband visible radiation from smooth planar metallic surfaces. In short, we do it with mirrors.

By pointing the camera at a mirror and rotating the mirror we provide a means of changing the viewpoint in a known way. Physically, the scene points are reflected by the moving mirror, and the resulting moving virtual scene is imaged in a stationary real camera, yielding image data which can be used to drive range recovery using "structure from known motion". Because the mirror's rotation axis is offset from the optic centre, rotation induces a translation of the scene relative to the optic centre. An equivalent but informative way of regarding the system draws on Fermat's principle, which indicates that one can consider the scene to be unreflected, but viewed by a virtual camera created by reflection in the mirror. As the mirror is moved, so this virtual camera moves, providing the multiple views. The mirror's rotation axis becomes the fixation point for stereo using multiple "virtual cameras".

As well as providing a compact mechanism, the device described recovers range over a wide field of view: apart from two blind directions, covering say 30°, this is panoramic. A further feature is that it is possible to recover range in a plane using only 1D image measurements; indeed this possibility is the one we explore experimentally here. Such a device may be useful for navigation.

The use of mirrors in not new in vision. They are of course used routinely in devices using active illumination (e.g. [5] describes an active IR rangefinder using a rotating mirror). In passive vision, plane mirrors were used by Cornog [2] in an early mechanism

BMVC 1992 doi:10.5244/C.6.4

for redirecting gaze, and are now used for the same purpose in commercially available tele-operated surveillance systems. A conical mirror [7] has been used to obtain a panoramic view of a scene, and by establishing correspondence with a second similar panoramic image taken from a different viewpoint, panoramic stereo recovery is possible. (A similar approach has been made using spherical projection with a fish-eye lens [6].)

We have, however, been unable to find a similar application of moving mirrors to range recovery. The closest analogue to our system is that of Ishiguro *et al* [3, 4], in which *real* cameras are moved in the same way that our *virtual* camera moves (see Figure 2b). These authors however analyse their results quite differently. They establish the range of a scene point from correspondence between just two points one from each of a panoramic stereo pair. In this work we recover each range value from several tens of image measurements, using a form of the spatio-temporal (or, in our case, spatio-angular) epipolar analysis of Bolles *et al* [1].

We describe and analyse the rotating mirror system in the following section. In Section 3, we show experiments using a implementation of the device using 1D image measurements to recover 2D scene information. The results are discussed in Section 4, along with a modification which removes blind directions, though at the expense of the 1D image solution.

2 Imaging using a mirror system

Figure 1 shows the arrangement of camera and mirror for the first device to be experimented with. The camera's optic axis defines $\hat{\mathbf{z}}$ and the optic centre is at the origin of

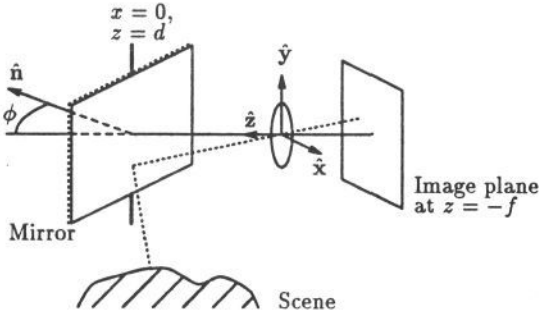

Figure 1: The rotating mirror device. The camera is placed in front of mirror which rotates about an axis parallel to $\hat{\mathbf{y}}$. Images are formed under perspective on the image plane at $z = -f$.

the physical camera's right-handed coordinate system $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$. Images are formed under perspective projection on the image plane at $z = -f$. The mirror is placed at a distance d along the optic axis, and rotates about the axis $(x = 0, z = d)$, parallel to $\hat{\mathbf{y}}$. The normal to the mirror is $\hat{\mathbf{n}}$, pointing into the mirror surface, so that when the mirror is rotated by angle ϕ as shown

$$\hat{\mathbf{n}} = (-\sin\phi, 0, \cos\phi)^T \ . \tag{1}$$

Figure 2 sketches the two ways of modelling the system outlined in the introductory section. Either (a) we consider the virtual scene imaged by the real camera, or (b) image the real scene in a virtual camera (b), where we note that in the latter case the coordinate system attached to the virtual camera has reversed parity due to reflection. The latter model makes clear that the virtual camera rotates about the mirror axis. Because this

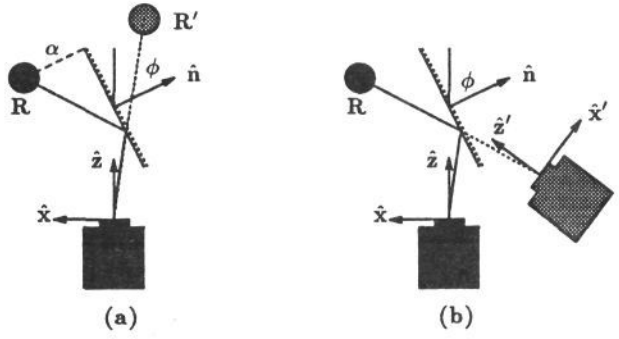

Figure 2: Two different ways of modelling the system. In (a) the virtual scene point (half filled) is imaged by the real camera (filled) and in (b) the real scene point is imaged in a "virtual camera".

axis does not pass through the real camera's optic centre, the optic centre of the virtual camera rotates *and* translates, the latter being necessary for depth recovery.

It is slightly simpler to derive the imaging conditions using model (a). Consider a scene point $\mathbf{R} = (X, Y, Z)^T$, its virtual reflection at $\mathbf{R}'$, and the image formed of it in the camera $\mathbf{r} = (x, y)^T$. Suppose that when the mirror is rotated by ϕ, the distance from $\mathbf{R}$ to the mirror is α, so that

$$\mathbf{R}_m = \mathbf{R} + \alpha\hat{\mathbf{n}} \tag{2}$$

is a point in the plane of the mirror. The equation of the mirror plane is

$$\mathbf{R}_m \cdot \hat{\mathbf{n}} = d\cos\phi\ , \tag{3}$$

whence

$$\alpha = d\cos\phi - \mathbf{R}\cdot\hat{\mathbf{n}}\ . \tag{4}$$

The virtual scene point is formed equidistant behind the mirror at

$$\mathbf{R}' = \mathbf{R} + 2\alpha\hat{\mathbf{n}} \tag{5}$$

$$= \begin{bmatrix} X\cos 2\phi + (Z-d)\sin 2\phi \\ Y \\ X\sin 2\phi - (Z-d)\cos 2\phi + d \end{bmatrix}\ . \tag{6}$$

Under perspective projection the virtual scene point is imaged at

$$\mathbf{r} = (x, y)^T = -f\mathbf{R}'/\mathbf{R}'\cdot\hat{\mathbf{z}}\ , \tag{7}$$

where f, the focal length, is assumed known from calibration. The image measurements are then of the form

$$\begin{aligned} m_x &= \frac{x}{f} = -\left(\frac{X\cos 2\phi + (Z-d)\sin 2\phi}{X\sin 2\phi - (Z-d)\cos 2\phi + d}\right) \\ m_y &= \frac{y}{f} = -\left(\frac{Y}{X\sin 2\phi - (Z-d)\cos 2\phi + d}\right)\ . \end{aligned} \tag{8}$$

Now consider what happens as the angle of rotation of the mirror ϕ is changed. The image of a particular scene point moves across the image, eventually leaving the field of view. By establishing correspondence, we construct the image locus $\mathbf{r}(\phi)$ of this point, from which we obtain a set of measurements $\{\ldots, m_{xi}, m_{yi}, \ldots\}$ where $m_{xi} = x(\phi_i)/f$ and $m_{yi} = y(\phi_i)/f$, all arising from the same scene point.

2.1 Recovering range

It is evident from equation (8) that the set of measurements provides an over constrained linear system for $\mathbf{R} = (X, Y, Z)^T$. In fact, it is most straightforward to recover

$$\mathbf{R}^* = (X^*, Y^*, Z^*)^T = (X, Y, Z - d)^T \ . \tag{9}$$

In other words, the natural place for the origin of coordinates is at the axis of rotation of the mirror, not the optic centre of the camera.

At a particular angle ϕ_i, equation 8 can be rewritten as

$$\begin{bmatrix} m_{xi}\sin 2\phi_i + \cos 2\phi_i & 0 & \sin 2\phi_i - m_{xi}\cos 2\phi_i \\ m_{yi}\sin 2\phi_i & 1 & -m_{yi}\cos 2\phi_i \end{bmatrix} \mathbf{R}^* = -d \begin{bmatrix} m_{xi} \\ m_{yi} \end{bmatrix} \ . \tag{10}$$

For k samples at different angles ϕ_i, k such matrices are blocked together to form

$$[\mathbf{A}]\mathbf{R}^* = \mathbf{b} \tag{11}$$

where $[\mathbf{A}]$ is an $2k \times 3$ matrix and $\mathbf{b}$ has length $2k$. Assuming independent measurements, with measurement i having weight W_{ii}, a least squares solution for the over-constrained system can be found by solving

$$[\mathbf{A}^T\mathbf{W}\mathbf{A}]\mathbf{R}^* = [\mathbf{A}^T\mathbf{W}]\mathbf{b} \ , \tag{12}$$

where $[\mathbf{A}^T\mathbf{W}\mathbf{A}]$ is a real symmetric 3×3 matrix, and $[\mathbf{W}]$ is the diagonal weight matrix.

Although this solution is straightforward enough, equation 8 indicates that we can recover depth without the m_y measurements using

$$\begin{bmatrix} m_{xi}\sin 2\phi_i + \cos 2\phi_i \\ \sin 2\phi_i - m_{xi}\cos 2\phi_i \end{bmatrix}^T \begin{bmatrix} X^* \\ Z^* \end{bmatrix} = -dm_{xi} \ . \tag{13}$$

This solution is especially useful provided we track features along the central horizontal raster, $y = 0$, from which which can recover range in the $Y = 0$ plane. For k measurements, equation (13) can be rewritten in terms of an $k \times 2$ matrix $[\mathbf{A}_{1D}]$ and length k vector $\mathbf{b}_{1D}$ analogous to $[\mathbf{A}]$ and $\mathbf{b}$, and the least squares solution is found by solving

$$[\mathbf{A}^T{}_{1D}\mathbf{W}\mathbf{A}_{1D}] \begin{bmatrix} X^* \\ Z^* \end{bmatrix} = [\mathbf{A}^T{}_{1D}\mathbf{W}]\mathbf{b}_{1D} \ . \tag{14}$$

Our experimental implementation of the rotating mirror system has pursued this recovery of 2D scene data from 1D image measurements in both simulation and using imagery. It is convenient to define 2D range, ρ, in a particular direction, γ, measured from the mirror's rotation axis as

$$\rho = \sqrt{X^{*2} + Z^{*2}} \quad \text{and} \quad \gamma = \tan^{-1}\left(\frac{+X^*}{-Z^*}\right) \tag{15}$$

respectively. This definition of γ relates simply to the definition of ϕ: if the mirror is set at angle ϕ, then the $x = 0$ vertical strip in the image in viewing in direction $\gamma = 2\phi$.

3 Experimental Results

As a preliminary test of noise sensitivity, range recovery was tested from artificially generated image contours. A typical trial is shown in Figure 3. One can imagine the device placed so that the mirror rotation axis is at the centre of a cross shaped room. At a number of directions γ around the device the actual (X^*, Z^*) values of the "walls" were used to simulate the locus that would be obtained on the image under mirror rotation from $\phi = -90°$ to $\phi = 90°$. Because the reflected ray rotates at twice the rate of the mirror rotation (the 2ϕ dependence of equations (8), this range covers the entire 360° field of view around the device. The field of view of the camera itself was limited to 40°. The image positions of the scene points were synthesized and then corrupted with Gauss random noise, and the least squares technique method described in Section 2 used to recover the values of X^* and Z^*, and hence the range and direction, ρ and γ.

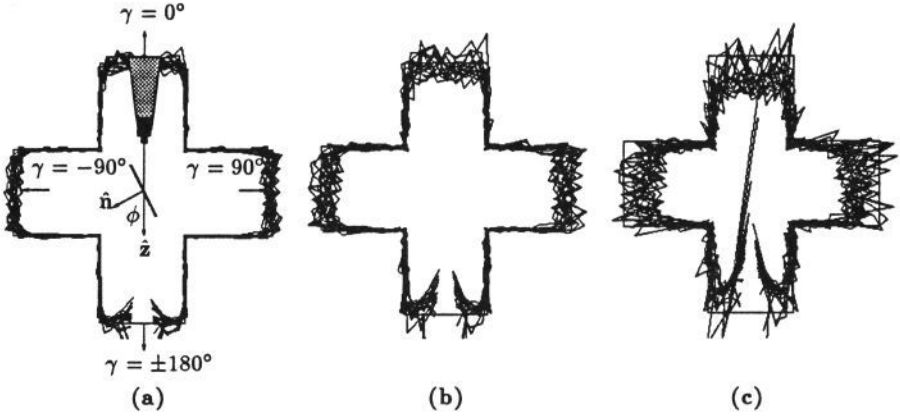

Figure 3: Simulations of range recovery using the rotating mirror under increasing error in image positions: (a) recovery with 2%, (b) with 4%, (c) with 8% noise added to locus positions.

Figure 3 shows the original outline of the "room" overlaid on the recovered range from several scans of the mirror for increasing values of image noise. The results show that range recovery is nearly panoramic. The shaded region indicates the invisible region around $\gamma = 0$ blocked out by the camera. The gaps ahead of the camera at $\gamma \approx \pm 180°$ occur when the mirror is edge on — that is, $\phi \approx \pm 90°$. In these positions, the m_{xi} are all close to zero and equation (13) degenerates to $X^* = 0$, leaving insufficient information to recover Z^*.

Another interesting feature is that as noise is introduced the range recovered tends to reduce, a useful conservative feature for navigation. The plots (especially (c)) suggest that this reduction in ρ is uniform in all directions (notwithstanding the impossibility of recovering depth when $\phi = 0$), and it is possible to show that the degradation is graceful. Suppose we assume that because of noise the measurements m_{xi} are actually independent of X^* and Z^*. Then equation 13 must be split into two parts, each independent of the m_{xi}. The parts are

$$m_{xi}\sin 2\phi_i X^* - m_{xi}\cos 2\phi_i Z^* = -d m_{xi} \quad (16)$$

$$\cos 2\phi_i X^* + \sin 2\phi_i Z^* = 0 \,. \quad (17)$$

Dividing out the first of these by m_{xi} and solving:

$$X^* = -d\sin 2\phi_i \quad Z^* = d\cos 2\phi_i \,, \quad (18)$$

whence

$$\rho = d \ , \tag{19}$$

independent of ϕ, as observed.

3.1 Experiments with imagery

As a prelude to building a dedicated device, we have exploited a robot arm to provide rotation.

The device is housed in a perspex safety cage in the corner of a visually cluttered laboratory, whose approximate plan view is shown in Figure 4. To establish scene points with known range, most of the inside of the perspex cage (A1 and A2) was "wallpapered" with black on white patterns, as shown in the Figure, although at (B) the device could see out. Objects were scattered about the walls (C1 and C2). For operational reasons, the wallpaper was place on the outside of the cage door at (D). These features are also marked on Figure 5, a panoramic view around the device. This image is for explanation only, and plays no part in the analysis.

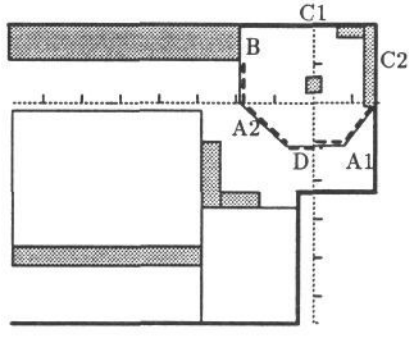

Figure 4: A plan of the workspace inside the cage. Thick lines are walls, the half shaded regions are populated with visually interesting equipment, and the lines are perspex safety cages. The dashed line represents the "wallpaper". The scale is in metres.

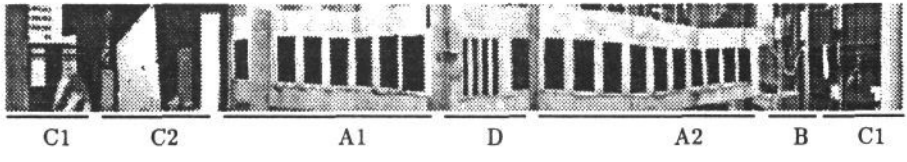

Figure 5: A panoramic view inside the cage. The labels refer to the text and Fig.4.

The camera and image capture electronics were calibrated to determine the optic centre relative to the top-left of the framestore ($x_c = 256.6(4)$ pixel, $y_c = 255.4(4)$ pixel) and focal length ($f_x = 1302(8)$pixel) and aspect ratio $s = 1.54(1)$, allowing framestore coordinates to be converted to world coordinates. The distance d between rotation centre and optic axis was determined as $0.176(1)m$. A single rotation scan was performed, moving the mirror angle ϕ between $-90°$ and $90°$ in $0.25°$ steps. Image rasters $-15 < y < +15$ centred about $y = 0$ (in world coordinates) were captured and edges derived using the Canny operator followed by hysteresis linking and thresholding, and the resulting edge information from the central $y = 0$ raster, consisting of $x-$position, orientation and contrast, stored. Near-horizontal edges were discarded as their intersection with the horizontal raster is likely to be uncertain.

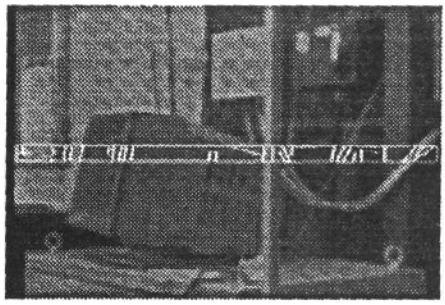

Figure 6: Edgels were computed in a central image block, and those of sufficient contrast linked into extended contours. The $x-$position, contrast and orientation of those in the central raster $y = 0$ were stored.

A simple matcher was used to link corresponding edgels up to form extended contours $x(\phi)$, using expected position, contrast and orientation as matching attributes. The change is $x-$position is bounded by $\Delta x = 0$ if the scene point is at range $\rho = 0$, and $\Delta x \approx f(2\Delta\phi)$ when the range is infinite. In our case this maximum was about 14 pixel. Of course, as the slope $dx/d\phi$ of a contour does not change markedly, the search range can be reduced after the first contour measurements are made. The angle of view of the camera itself was around 23°, and the maximum number of matches in a contour was thus 46. We retained (somewhat arbitrarily) all those with greater than 30 matches for further analysis. The contours $x(\phi)$ are shown in Figure 7.

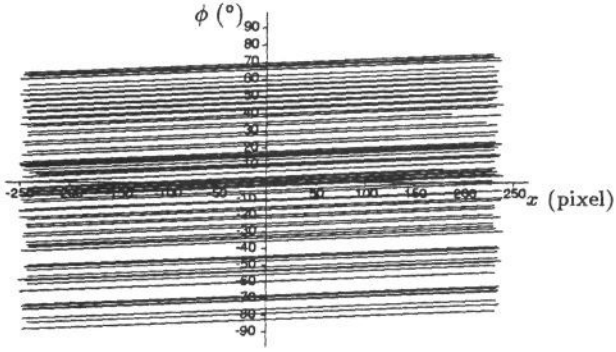

Figure 7: Contours $x(\phi)$ created by matching edgels in successive frames. The x-axis is in pixels, the ϕ-axis in degrees.

The set of discrete measurements from each contour was analysed using the least squares method of Section 2, and the recover scene points plotted in Figure 8. The outline of the cage is recovered well, as are objects from the side walls. Note that the depths where the device could see out of the cage are indeed greater. The two points recovered at $Z^* \approx -5m$ caused some surprise. However, as we noted earlier, the wallpaper on the cage door was on the *outside* and these points correspond to some strong reflections of the equipment at A in the perspex cage.

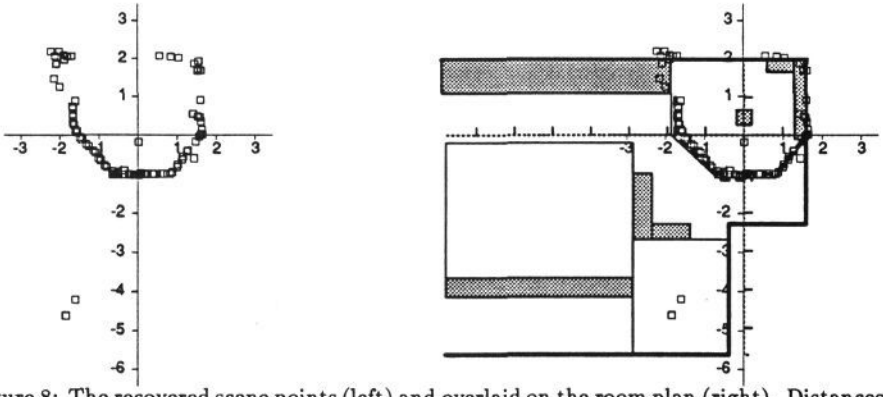

Figure 8: The recovered scene points (left) and overlaid on the room plan (right). Distances are in metres.

In a further experiment, the "wallpaper" was removed, and three objects (polystyrene mannequins) introduced closer to the device. The panorama (Figure 9) shows that much of the scene is visually cluttered, and we again expect to suffer problems of reflections in the cage walls, making it difficult to interpret the results in terms of specific objects. Nonetheless, the back wall is now interpretable, and all the depths are broadly as expected (Figure 10).

Figure 9: A panoramic view of the entire laboratory.

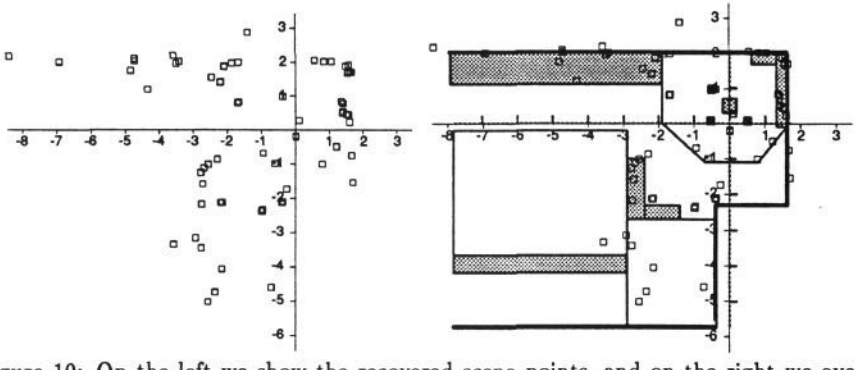

Figure 10: On the left we show the recovered scene points, and on the right we overlay them on the room plan for comparison. The scale is in metres.

4 Discussion

The device described shows considerable promise as a compact, wide angle of view passive vision sensor, and seems most applicable to autonomous navigation. We now discuss some of the perceived drawbacks and merits of the system.

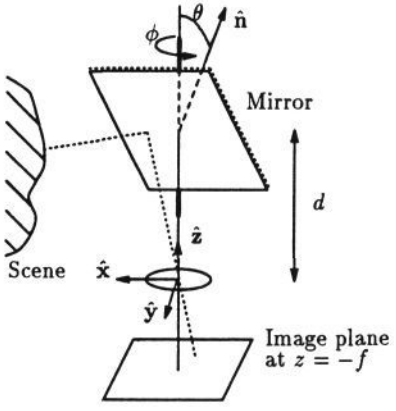

Figure 11: The second system. The camera is placed with its optic axis vertical and the mirror tilted by θ ($\theta = 45°$ is the likely choice). The mirror rotates about the camera's optic axis $\hat{\mathbf{z}}$.

First it is appropriate to compare the merits of our analysis with that in Ishiguro *et al* [3, 4] in their analagous real camera system. As their camera rotates, by taking two *vertical* strips of one pixel width from each image they build up a panoramic stereo pair of images. By establishing correspondence between pairs of features in the two images they are able to compute range. Essentially then they must ensure that a feature observed in one pixel-wide strip is captured in the other, and so they must rotate the camera in angular increments of approximately $1/f$ radian, where f is the focal length in pixels. As $f \approx 1000$ this requires high angular resolution. Our approach is to establish correspondence and track features through each successive image at much lower angular resolution. In the experimental system that recovers range in the 2D plane, this involves taking a single *horizontal* raster from each image. By doing this we can rotate by larger angles, but still recover several tens of image measurements from which to recover the range of single scene point, rather than just two measurements.

One disadvantage of the mirror system compared with rotating a real camera with the configuration described is that a panoramic view is marred by blind spots when the mirror is edge-on and when the camera looks at itself. A different configuration which remedies this problem is shown in Figure 11. The fixed camera is positioned with its optic axis $\hat{\mathbf{z}}$ vertical and the mirror is tilted at some fixed angle θ (likely to be 45°) and rotates through angles ϕ about the optic axis. The analysis closely follows that of Section 2, but the normal to the mirror is now

$$\hat{\mathbf{n}} = (\sin\theta\cos\phi,\ \sin\theta\sin\phi,\ \cos\theta)^T\ . \tag{20}$$

The perpendicular distance to the mirror plane is now fixed at $d\cos\theta$, and the virtual scene point is therefore at

$$\mathbf{R}' = \mathbf{R} + 2(d\cos\theta - \mathbf{R}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}} \tag{21}$$

The measurements are

$$m_{xi} = \frac{x(\phi_i)}{f} = -\frac{X'(\phi_i)}{Z'(\phi_i)} \tag{22}$$

$$m_{yi} = \frac{y(\phi_i)}{f} = -\frac{Y'(\phi_i)}{Z'(\phi_i)}\ , \tag{23}$$

and as in Section 2 these can be used as a set of linear equations for $\mathbf{R}^*$:

$$[\mathbf{A}]\mathbf{R}^* = \mathbf{b} \ . \tag{24}$$

where for k measurements $[\mathbf{A}]$ is an $2k \times 3$ matrix and $\mathbf{b}$ has length $2k$. The contribution from the ith measurement to $[\mathbf{A}]$ is

$$[\mathbf{A}] = \begin{bmatrix} \vdots & \vdots & \vdots \\ (2\mathrm{S}_\theta^2\mathrm{C}_{\phi_i}^2 - 1 + m_{xi}\mathrm{S}_{2\theta}\mathrm{C}_{\phi_i}) & (\mathrm{S}_\theta^2\mathrm{S}_{2\phi_i} + m_{xi}\mathrm{S}_{2\theta}\mathrm{S}_{\phi_i}) & (\mathrm{S}_{2\theta}\mathrm{C}_{\phi_i} + m_{xi}\mathrm{C}_{2\theta}) \\ (\mathrm{S}_\theta^2\mathrm{S}_{2\phi_i} + m_{yi}\mathrm{S}_{2\theta}\mathrm{C}_{\phi_i}) & (2\mathrm{S}_\theta^2\mathrm{S}_{\phi_i}^2 - 1 + m_{yi}\mathrm{S}_{2\theta}\mathrm{S}_{\phi_i}) & (\mathrm{S}_{2\theta}\mathrm{S}_{\phi_i} + m_{yi}\mathrm{C}_{2\theta}) \\ \vdots & \vdots & \vdots \end{bmatrix} \tag{25}$$

and to $\mathbf{b}$ is

$$\mathbf{b} = d \begin{bmatrix} \vdots \\ m_{xi} \\ m_{yi} \\ \vdots \end{bmatrix} \ , \tag{26}$$

where $\mathrm{S}_\theta^2 = \sin^2\theta$, $\mathrm{S}_{2\theta} = \sin 2\theta$ and so on. The obvious disadvantage with this modified system is that there is no 1D version available using matching within a single raster. The image will rotate about the optic centre, and tracking would have to be performed in 2D, using say a corner detector.

Another difficulty with the present treatment of the data is that a sparse depth map is recovered. Of course this is inevitable if features invariant to raw intensity changes are sought. However, as the viewpoint changes are small, it may be that a matching process using attributes closer to the raw intensity would suffice. The technique that immediately suggests itself is dynamic time warping. This technique might also address a further and most pressing problem with the present arrangement, that of speed. For the device to be practical, we must complete a complete scan in say a couple of seconds. Using $\Delta\phi = 0.25°$ then requires an acquisition and processing time per frame of order 3 *ms*. This suggests that we should consider only the 1D device as feasible at present and use fast linear sensor arrays with dedicated processing hardware.

References

[1] R C Bolles, H H Baker, and D H Marimont. Epipolar-plane image anlysis: an approach to determining structure from motion. *International Journal of Computer Vision*, 1:7–55, 1987.

[2] K H Cornog. Smooth pursuit and fixation for robot vision. Master's thesis, Department of Electrical Engineering and Computer Science, MIT, 1985.

[3] H Ishiguro, M Yamamoto, and S Tsuji. Omni-direction stereo for making global map. In *3rd International Conference on Computer Vision, Osaka, 1990*, pages 540–547, Washington DC, 1990. IEEE Computer Society Press.

[4] H Ishiguro, M Yamamoto, and S Tsuji. Omni-directional stereo. *IEEE Transactions on PAMI*, 14(2):257–262, 1992.

[5] G L Miller and E R Wagner. An optical rangefinder for autonomous cart navigation. In *Proceedings SPIE Mobile Robots II*, volume 852, pages 132–144, Bellingham, Washington, 1987. SPIE.

[6] T Morita. Measurement in three dimensions by motion stereo and spherical mapping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1989*, pages 422–428, Washington DC, 1989. IEEE Computer Society Press.

[7] Y Yagi and S Kawato. Panoramic scene analysis with conic projection. In *IEEE International Workshop on Intelligent Robots and Systems*, pages 181–190, Washington DC, 1990. IEEE Computer Society Press.