

Real-Time Omnidirectional and Panoramic Stereo *

Joshua Gluckman, Shree K. Nayar and Keith J. Thoresz

Department of Computer Science

Columbia University

New York, NY 10027

Abstract

Traditional stereo systems have a small field of view which limits their usefulness for certain applications. By imaging curved mirrors, the field of view of conventional cameras can be enhanced. In this paper, we present the design of a compact panoramic stereo camera which uses parabolic mirrors and is capable of producing 360° panoramic depth maps at the rate of several frames per second. Video surveillance, autonomous navigation, and site modeling are some of the applications which will benefit from such a sensor.

1 Motivation and Background Work

Automated video surveillance systems need the ability to detect and track movements over a large 3-D space. Similarly, autonomous navigation systems must perceive and avoid obstacles which may not be in the field of view of conventional cameras. Site modeling, the recovery of 3-D structure of a large scene, is another application which typically requires range data over a large field of view. In these applications, multiple cameras, moving parts or active sensors are often used to compensate for the small field of view of traditional stereo systems. We have designed a stereo system which uses a combination of lenses and curved mirrors (known as a *catadioptric* system) to capture a wide field of view without moving parts.

Previously, several researchers have used panoramic images to compute depth maps. Ishiguro *et al.* [1992] used panoramic images obtained from a rotating camera to compute depth. Similar rotating systems have been proposed by [Murray, 1995], [McMillan and Bishop, 1995], [Benosman *et al.*, 1996] and [Kang and Szeliski, 1997]. The disadvantage of such systems is that they require mechanical scanning and are thus not practical for real-time applications such as surveillance and navigation.

Panoramic images can be acquired in real-time by imaging curved, mirrored surfaces. One of the first uses of

mirrors for stereo was in [Nayar, 1988], where Nayar suggested a wide field of view stereo system consisting of a conventional camera pointed at two specular spheres. A similar system using two convex mirrors, one placed on top of the other, was proposed by Southwell *et al.* [Southwell *et al.*, 1996]. However, in both these systems the projection of the scene produced by the curved mirrors is *not* from a single viewpoint. Violation of the “single viewpoint assumption” implies that the pinhole camera model can not be used, thus making calibration and correspondence a more difficult task.

In [Nayar and Baker, 1997], Nayar and Baker derived the class of catadioptric cameras which do produce a single view point. Later, Nene and Nayar [1998] presented several different stereo configurations using planar, parabolic, elliptic, and hyperbolic mirrors which ensure a single viewpoint. A catadioptric stereo system using hyperbolic mirrors was implemented by Chaen *et al.* [1997]. In this paper, we present the design of a compact panoramic stereo sensor using parabolic mirrors, which is capable of producing panoramic depth maps in real-time.

2 Panoramic Stereo Camera

Figure 1 is a diagram of our compact panoramic stereo sensor, which is composed of two omnidirectional cameras, aligned vertically (coaxial). Each omnidirectional camera consists of a conventional camera, telecentric optics, and a parabolic mirror. Telecentric optics approximates orthographic projection which, together with a parabolic mirror, form a wide-angle imaging system with a single center of projection located at the focal point of the parabola [Nayar, 1997]. The reasons for aligning the sensors vertically are twofold. First, by placing one sensor above the other the singularities (where depth cannot be computed near the epipoles) occur where the cameras occlude themselves. Second, the panoramic images have corresponding, parallel epipolar lines, thus lending themselves to real-time stereo processing.

A single camera could be used by placing one smaller

*This work was supported in part by DARPA's VSAM Image Understanding Program, under ONR contract N00014-97-1-0553.

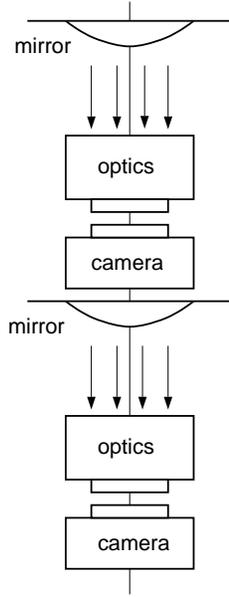


Figure 1: Schematic diagram of a compact panoramic stereo camera. Two omnidirectional cameras are used to capture panoramic stereo data. Omnidirectional imaging is achieved by the orthographic projection of a parabolic reflecting surface. Telecentric optics is used to achieve orthographic projection while keeping the camera close to the mirror.

mirror on top of the other as in [Southwell *et al.*, 1996]. However, the resolution of the two panoramas will be significantly different which leads to less robust stereo matching.

2.1 Field of view

The field of view (FOV) of each omnidirectional camera is determined by the extent of the parabola. If the parabola extends beyond its focal point, it is possible to obtain a FOV greater than a hemisphere. Each omnidirectional camera also produces a cone of occlusion, where the camera sees itself in the mirror. Figure 2 shows the field of view shared by the two cameras, where depth may be computed. When the cameras are coaxial the shared FOV is the FOV of the bottom camera minus the occlusion cone of the top omnidirectional camera.

2.2 Generating Panoramas

We are able to generate full 360° panoramas by projecting the input image onto a cylinder. As shown in figure 3, the mapping is done by first orthographically projecting the image point \mathbf{p} onto the parabola, and then projecting to the point \mathbf{p}' by a ray passing through the focal point of the parabola.

When generating a panorama, we back project each pixel in the panoramic image to a point in the input image and use bilinear interpolation to find the intensity at that point. Using the following equation for a parabola (coor-

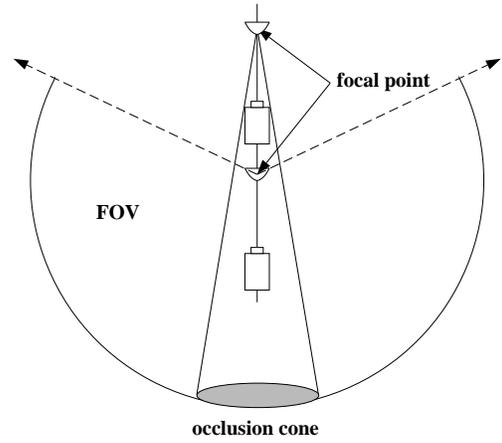


Figure 2: The field of view. By extending the parabola beyond the focal point, parabolic mirrors are capable of capturing a field of view greater than a hemisphere. An occlusion cone is created where the camera “sees itself.”

dinate system located at the focal point),

$$z = \frac{r^2 - x^2}{2r}, \quad (1)$$

the back projection function which relates height z' in the panorama, to distance x from image center becomes

$$z' = \frac{R(r^2 - x^2)}{2rx}. \quad (2)$$

The radii of the parabola and cylinder are denoted by r and R , both measured in pixels. While equation (2) only defines one dimension in the panorama the other dimension is simply a function of polar angle in the image. Next, we discuss the epipolar geometry of both the captured image and the panoramic image.

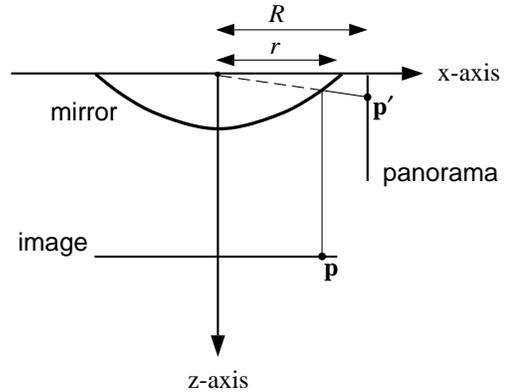


Figure 3: Mapping an image point to a panorama. A point in the image \mathbf{p} is mapped to the point \mathbf{p}' in the panorama.

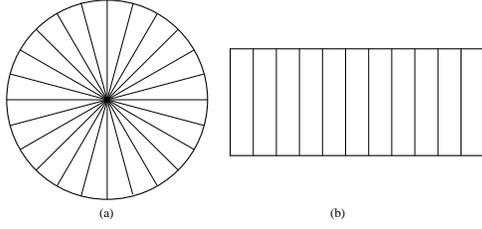


Figure 4: Epipolar geometry. (a) The epipolar lines for a pair of coaxial omnidirectional images are radial lines. (b) When the images are projected onto a panorama the epipolar lines become parallel.

2.3 Epipoar Constraint

An important part of any stereo system is the epipolar constraint. This constraint reduces the problem of finding corresponding points to a 1-D search. The epipolar constraint for catadioptric systems has been studied by [Nene and Nayar, 1998] and [Svoboda *et al.*, 1998]. For parabolic mirrors, corresponding points must lie on conics (epipolar curves). However, when the parabolas are vertically aligned such that the vertices and focal points are coaxial, the curves reduce to radial lines (see figure 4). Once the image of the parabola is projected onto a cylinder (panoramic image) the epipolar lines become parallel. If each image in the stereo pair is projected onto a cylinder of the same size, the epipolar lines will match up. This implies that R in equation (2) should be the same for the projection function of each stereo image in order to ensure matched epipolar lines. Matched epipolar lines are desirable because efficient search methods can be used to perform stereo correspondence.

2.4 Computation of Depth

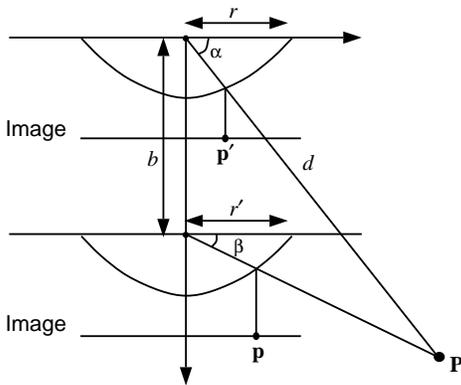


Figure 5: Triangulation and the computation of depth. Given a pair of image correspondences \mathbf{p} and \mathbf{p}' , the scene point \mathbf{P} is found by triangulation.

Once correspondence between image points has been established, depth computation by triangulation is straightforward. As shown in figure 5, each corresponding im-

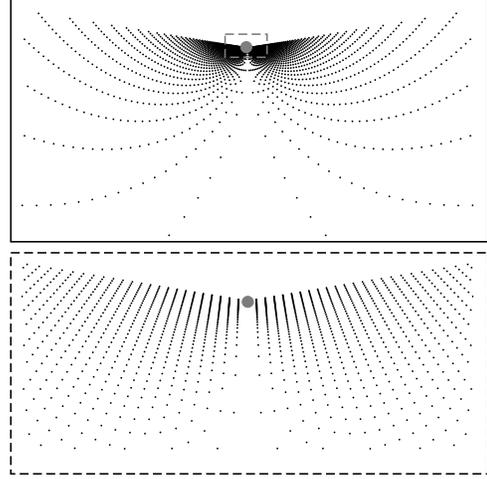


Figure 6: Depth resolution. Each point represents the depth of a pair of corresponding image points taken from a single, radial epipolar line. The bottom graph, which is an enlargement, shows that the depth resolution becomes uniform close to the parabola. The two grey dots represent the location of the bottom parabola in the stereo system.

age point is the orthographic projection of a point on the parabola, which can be defined by the angles α and β . From the law of sines, depth d is computed as

$$d = \frac{\cos(\beta)}{\sin(\alpha - \beta)} b, \quad (3)$$

where b is the baseline of the stereo system (see figure 5). Letting x and x' be the distances of the image points \mathbf{p} and \mathbf{p}' from the center of the parabolas in the image, then the angles α and β are determined by,

$$\tan(\alpha) = \frac{r^2 - x^2}{2rx} \quad (4)$$

and

$$\tan(\beta) = \frac{r'^2 - x'^2}{2r'x'}, \quad (5)$$

where r and r' are the radii of the parabolas. Note that in contrast to traditional stereo, depth computation is independent of the focal length of the camera (due to orthographic projection).

Alternatively, distance can be defined as a function of disparity in the panoramic image. If x and x' are the heights of corresponding points in the top and bottom panoramas, disparity $x - x'$ is inversely related to depth as

$$\left(\frac{x - x'}{\sqrt{R^2 + x^2}}\right)b = \frac{1}{d}. \quad (6)$$

2.5 Depth Resolution

Because of the non-uniform spatial resolution introduced by using curved mirrors, depth resolution is more

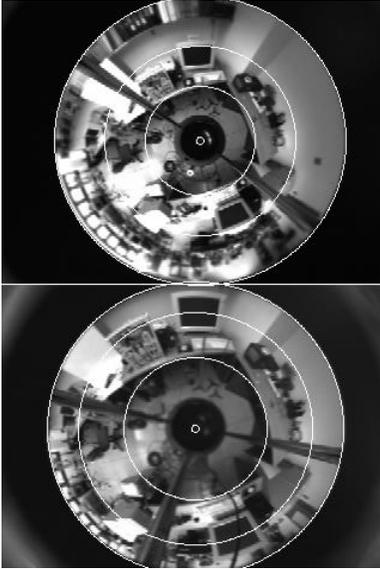


Figure 7: User interface of the calibration tool. The stereo omnidirectional images are calibrated using the radius and center of the parabola in each image, which are found by fitting a circle to the image of the parabola.

complicated for omnidirectional stereo than traditional stereo. Figure 6 shows a sampling of the depth resolution in 2-D. The sampling is obtained by computing depth d for every possible pair of image correspondences x and x' using the following equation

$$d = \frac{(r^2 + x^2)x'}{(r^2 - x^2)x' - (r^2 - x'^2)x} b, \quad (7)$$

which is derived from equations (3), (4), and (5) (assuming equal radii of the two parabolas). Note that the baseline b effects the scale and the radius r effects the density of the graph in figure 6. While equation (7) only describes depth in 2-D, the depth resolution in 3-D is further complicated by the fact that it is dependent on the volume of intersection of the two solid angles subtended by the corresponding image points. Finding the baseline which optimizes the depth resolution for a particular scene is an interesting issue for future study.

2.6 Calibration

Before panoramas can be generated and depth computed, the system must be calibrated. Due to different focal settings between the cameras, the sizes of the parabolas in the image may differ. The center of the parabola may not be located at the center of the image. Furthermore, a slight rotation about the vertical axis may exist between the two cameras. Therefore, we need to determine the parabola radii and centers as well as the rotational shift between the two cameras.

We have implemented an interactive tool to find these calibration parameters. As shown in figure 7 a circle is



Figure 8: A compact panoramic stereo camera. In this design we used parabolas with a $360^\circ \times 200^\circ$ field of view.

fit to each parabolic image. This circle determines the center and radius of the parabola. Of course, depending on the extent of the parabola the measured radius may not be r the radius at the focal point, which is used in equations (1), (3), (4), (5), and (7). However, in practice the extent of the parabola is known precisely and r may be estimated from the radius in the image r_i using

$$r = r_i(\sec \phi - \tan \phi), \quad (8)$$

where ϕ is the angular extent of the parabola beyond the focal point.

Once r is obtained, the images can be projected onto a panoramic image, where the epipolar lines are parallel. Then, by manually selecting a few corresponding points the rotational shift between the images can be measured and removed. At this point, the epipolar lines will match up and we can proceed with scanline stereo matching.

An additional feature of the calibration tool is an adjustable field of view. Because of the computational demands of stereo matching, it is often the case that only a portion of the image can be used (alternatively resolution can be reduced). For this reason, we allow the user to define a portion (a field of view) within each image which defines the active part of the stereo search. Note that the panorama will always cover 360° ; the field of view adjustment will only effect the elevation of the panorama. In figure 7 the bounds of this area are shown as circles defined by the user.

3 Real-time Implementation

We have implemented a real-time stereo algorithm which processes images taken by the panoramic stereo sensor (shown in figure 8). After digitization, each image is projected onto a cylindrical (panoramic) image. The radius R of the cylinder is set to the average of the two radii



Figure 9: Panoramic images and disparity map. The disparity map (on top) was computed from the two panoramic images (on bottom).

of the parabolas. Because r , r' and R , the parameters of the projection function, are fixed at run-time the warping function can be computed off-line and implemented as a lookup table. At run-time, each pixel in the panoramic image is back projected through the lookup table to a point in the input image. Bilinear interpolation, implemented in integer arithmetic for speed, is used to improve warping quality. Figure 9 (on bottom) shows an example panoramic stereo pair.

Once projected onto a cylinder, stereo matches are found using a standard window-based correlation search with a sum of absolute differences correlation measure. Because the epipolar lines are matched and parallel in the panoramic images, this search is implemented efficiently as in [Faugeras *et al.*, 1993]. Further speed enhancements are made using the MMX SIMD (single instruction multiple data) instruction set. An example panoramic disparity image is shown in figure 9 (on top).

For panoramic images of 600×60 pixels and a depth search of 32 pixels, we have achieved a throughput of approximately 7 frames per second on a standard 300 Mhz Pentium II PC. The image size of 600×60 was chosen because in panoramic stereo we are primarily interested in horizontal resolution.

References

- [Benosman *et al.*, 1996] R. Benosman, T. Maniere, and J. Devars. Multidirectional stereovision sensor, calibration and scene reconstruction. In *Proceedings of the Int'l Conference on Pattern Recognition*, 1996.
- [Chaen *et al.*, 1997] A. Chaen, K. Yamazawa, N. Yokoya, and H. Takemura. Omnidirectional stereo vision using hyperomni vision. Technical Report 96-122, IEICE, Feb. 1997. in Japanese.
- [Faugeras *et al.*, 1993] O. Faugeras, B. Hotz, H. Mathieu, T. Vieville, Z. Zhang, P. Fau, E. Theron, L. Moll, G. Berry, J. Vuillemin, P. Bertin, and C. Proy. Real-time correlation-based stereo: algorithm, implementation and application. Technical Report 2013, INRIA Sophia Antipolis, 1993.
- [Ishiguro *et al.*, 1992] H. Ishiguro, M. Yamamoto, and S. Tsuji. Omnidirectional stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:257–262, 1992.
- [Kang and Szeliski, 1997] S.B. Kang and R. Szeliski. 3-d scene data recovery using omnidirectional multibaseline stereo. *International Journal of Computer Vision*, 25(2), November 1997.
- [McMillan and Bishop, 1995] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. In *Computer Graphics (SIGGRAPH '95)*, pages 39–46, 1995.
- [Murray, 1995] D.W. Murray. Recovering range using virtual multicamera stereo. *Computer Vision and Image Understanding*, 61(2):285–291, 1995.
- [Nayar and Baker, 1997] S.K. Nayar and S. Baker. Catadioptric image formation. In *Proceedings of the 1997 DARPA Image Understanding Workshop*, May 1997.
- [Nayar, 1988] S.K. Nayar. Robotic vision system. United States Patent 4,893,183, Aug. 1988.
- [Nayar, 1997] S.K. Nayar. Omnidirectional video camera. In *Proceedings of the 1997 DARPA Image Understanding Workshop*, May 1997.
- [Nene and Nayar, 1998] S.A. Nene and S.K. Nayar. Stereo with mirrors. In *Proceedings of the 6th International Conference on Computer Vision*, Bombay, India, January 1998. IEEE Computer Society.
- [Southwell *et al.*, 1996] D. Southwell, A. Basu, M. Fiala, and J. Reyda. Panoramic stereo. In *Proceedings of the Int'l Conference on Pattern Recognition*, 1996.
- [Svoboda *et al.*, 1998] T. Svoboda, T. Pajdla, and V. Hlavac. Epipolar geometry for panoramic cameras. In *Fifth European Conference on Computer Vision*, pages 218–232, 1998.