A Real-Time Catadioptric Stereo System Using Planar Mirrors *

Joshua Gluckman and Shree K. Nayar
Department of Computer Science
Columbia University
New York, NY 10027

Abstract

By using mirror reflections of a scene, stereo images can be captured with a single camera. Single camera stereo provides both geometric and radiometric advantages over traditional two camera stereo. In this paper, we discuss these advantages and show that the epipolar geometry is restricted to the class of planar motions. In addition we have implemented a real-time system which demonstrates the viability of stereo with mirrors as an alternative to traditional two camera stereo.

1 Introduction

Optical systems consisting of a combination of refracting (lens) and reflecting (mirror) elements are called *catadioptric* systems [Hecht and Zajac, 1974]. By using two or more mirrored surfaces, multiple views of a scene can be captured by a single camera (catadioptric stereo). Single camera stereo provides several advantages over traditional two camera stereo.

- **Identical System Parameters:** Lens, CCD and digitizer parameters such as blurring, lens distortions, focal length, spectral responses, gain, offset, pixel size, etc. are identical for the stereo pair (assuming ideal mirrors). Having identical system parameters minimizes the differences between the two views, thus facilitating robust stereo matching.
- **Ease of Calibration:** Because only a single camera and digitizer is used, there is only one set of intrinsic calibration parameters. As we will show, the extrinsic calibration parameters are constrained by planar motion. Together these constraints reduce the total number of calibration parameters from 16 to 10.
- **Data Acquisition:** Camera synchronization is not an issue because only a single camera is used. Stereo data can easily be acquired and conveniently stored with a standard video recorder without the need to synchronize multiple cameras.

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

With these advantages in mind, we present the design and implementation of a real-time stereo system which uses only a single camera and two planar mirrors. In addition, we analyze the geometry and calibration of stereo with planar mirrors in an arbitrary configuration and show that the epipolar geometry is restricted to planar motion. The planar motion constraint implies that the fundamental matrix for stereo with planar mirrors depends upon 6 parameters instead of 7 for traditional stereo.

Previously, several researchers have demonstrated the use of both curved and planar mirrors to acquire stereo data. For a discussion of panoramic stereo with curved mirrors see [Gluckman *et al.*, 1998] (in these proceedings). Goshtasby and Gruver [1993] designed a single camera stereo system using a pair of planar mirrors connected by a hinge. Mathieu and Devernay [1993] and Inaba *et al.* [1993] used four planar mirrors to create two virtual cameras with vergence controlled by the angle between two of the mirrors. In contrast to these, our system does not require the mirrors to be in a specific configuration. In addition we have implemented a real-time system which demonstrates robustness stereo matching when only a single camera is used.

Previous *real-time* stereo systems have used two or more cameras [Faugeras *et al.*, 1993] [Matthies, 1993] [Kanade *et al.*, 1996] [Konolige, 1997]. Because more than one camera is used, the images must be processed in order to compensate for the differences in camera response either by applying the Laplacian of the Gaussian or by using normalized correlation. These steps, which can be ignored in single camera stereo, are both computationaly intensive and result in loss of information. In the following section we derive the geometry of a catadioptric system with a single camera and two planar mirrors in an arbitrary configuration.

2 Geometry and Calibration

Previously, researchers have looked at the geometry of catadioptric systems in calibrated settings, where the mirrors are placed in specific configurations [Goshtasby

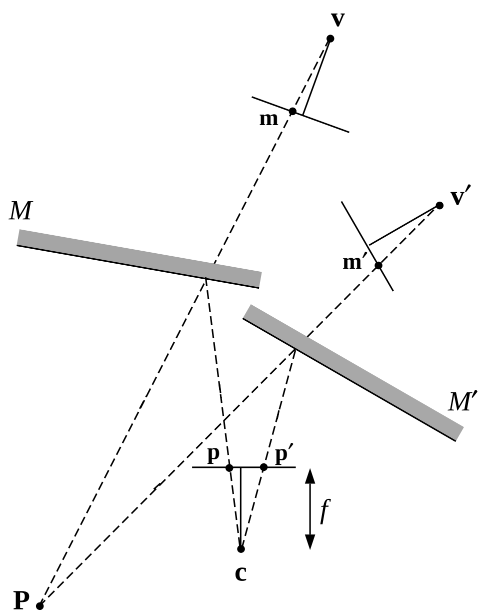

Figure 1: Stereo image formation with a single camera and two planar mirrors. A scene point $\mathbf{P}$ reflected off mirrors M and M' is imaged as if seen from two different viewpoints $\mathbf{v}$ and $\mathbf{v}'$.

and Gruver, 1993], [Inaba *et al.*, 1993] and [Mathieu and Devernay, 1993]. Here, we analyze the geometry of two mirrors and a single camera with the mirrors placed in an arbitrary configuration.

Figure 1 depicts the geometry of a catadioptric system with two planar mirrors. A scene point $\mathbf{P}$ is imaged as if seen from two different viewpoints $\mathbf{v}$ and $\mathbf{v}'$. The location of the two virtual pinholes is found by reflecting the camera pinhole about each mirror. Reflecting the optical axis of the camera about the mirrors determines the optical axes and thus the orientations of the two virtual cameras. The virtual image planes exist at a distance f, the focal length of the camera, along the optical axes of the two virtual cameras. Therefore, the locations and orientations of the two virtual cameras are determined by the orientations and distances of the two mirrors with respect to the pinhole and optical axis of the camera.

2.1 Relative Orientation

In traditional stereo with two cameras there are no quantitative restrictions on the relative orientation between the two cameras. However, constraints do exist for the two virtual cameras created when two planar mirrors are imaged by a single camera. It turns out that the relative orientation is restricted to planar motion (the direction of translation must lie in the plane normal to the axis of rotation). This constraint reduces the number of degrees of freedom of relative orientation from 6 to 5 (3 for rotation and 2 for translation in a plane).

To derive this result we consider the relative orientation between the two reflected viewpoints $\mathbf{v}$ and $\mathbf{v}'$. Each virtual viewpoint is related to the camera center by the following equations:

$$\mathbf{v} = \mathbf{D}_1\mathbf{c} \tag{1}$$

and

$$\mathbf{v}' = \mathbf{D}_2\mathbf{c}, \tag{2}$$

where $\mathbf{D}_1$ and $\mathbf{D}_2$ are reflection transformations. Then the relative orientation $\mathbf{D}$ becomes,

$$\mathbf{D} = \mathbf{D}_2\mathbf{D}_1^{-1}. \tag{3}$$

Representing the two mirrors as planes with normals $\mathbf{n}_1$ and $\mathbf{n}_2$ and distances d_1 and d_2 measured from $\mathbf{c}$ the camera center, the reflection transformations for the two mirrors are given by

$$\mathbf{D}_1 = \begin{bmatrix} \mathbf{I} - 2\mathbf{n}_1\mathbf{n}_1^T & 2d_1\mathbf{n}_1 \\ \mathbf{0} & 1 \end{bmatrix} \tag{4}$$

and

$$\mathbf{D}_2 = \begin{bmatrix} \mathbf{I} - 2\mathbf{n}_2\mathbf{n}_2^T & 2d_2\mathbf{n}_2 \\ \mathbf{0} & 1 \end{bmatrix}. \tag{5}$$

Because the inverse of a reflection transformation is itself, the relative orientation of the two virtual cameras is simply,

$$\mathbf{D} = \mathbf{D}_2\mathbf{D}_1 = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \tag{6}$$

where

$$\mathbf{R} = \mathbf{I} + 4(\mathbf{n}_1 \cdot \mathbf{n}_2)\mathbf{n}_1\mathbf{n}_2^T - 2\mathbf{n}_1\mathbf{n}_1^T - 2\mathbf{n}_2\mathbf{n}_2^T, \tag{7}$$

and

$$\mathbf{t} = 2d_1\mathbf{n}_1 - (2d_1(\mathbf{n}_1 \cdot \mathbf{n}_2) + 2d_2)\mathbf{n}_2. \tag{8}$$

The rotation matrix $\mathbf{R}$ has a rotational axis of $\mathbf{n}_1 \times \mathbf{n}_2$ and from (8) the direction of translation lies in the plane defined by $\mathbf{n}_1$ and $\mathbf{n}_2$. Therefore, the rotational axis is normal to the plane containing the direction of translation (planar motion).

Planar motion has been studied in the context of mobile robotics [Beardsley and Zisserman, 1995], where motion over a ground plane is modeled by planar motion. For such scenarios Vievill and Lingrand [1995] and Armstrong *et al.* [1996] have used planar motion to help constrain the self-calibration problem.

As we have seen, single camera stereo with two planar mirrors constrains the external calibration parameters to planar motion. Because only a single camera is used, the intrinsic parameters (focal length, pixel size, image center, skew) are exactly the same for the two stereo views. Together, these constraints place restrictions on the epipolar geometry.

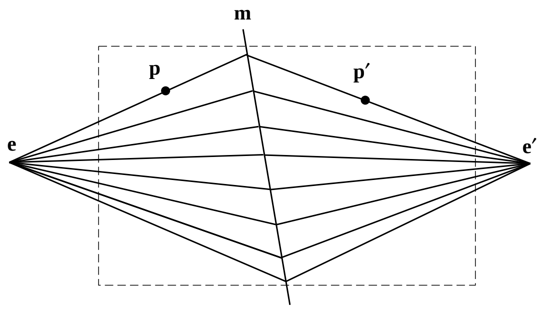

Figure 2: The epipolar geometry of planar motion. When motion is constrained to lie in a plane, all corresponding epipolar lines must intersect at $\mathbf{m}$ the image of the axis of rotation. Therefore, the two epipoles $\mathbf{e}$ and $\mathbf{e}'$ and the line $\mathbf{m}$ completely determine the epipolar geometry.

2.2 Epipolar Geometry

One way to describe planar motion between a pair of cameras is by a rotation about one of the camera centers and a translation in a direction normal to the axis of rotation. Alternatively, planar motion can be represented by a pure rotation of one of the cameras about an axis not necessarily passing through the camera center (called the *screw axis*). When the internal calibration of the two cameras is identical, the image projection of the screw axis is the same for both cameras. Therefore, corresponding epipolar lines must intersect on the line which is the image projection of the screw axis.

As shown in figure 2, the epipolar line of a point $\mathbf{p}$ is the line containing epipole $\mathbf{e}'$ and the intersection of the image of the screw axis $\mathbf{m}$ with the line through epipole $\mathbf{e}$ and point $\mathbf{p}$. If $\mathbf{p}$ and $\mathbf{p}'$ are corresponding points then

$$\mathbf{p}' \cdot (\mathbf{e}' \times (\mathbf{m} \times (\mathbf{e} \times \mathbf{p}))) = \mathbf{0}, \tag{9}$$

which implies that the fundamental matrix has the form

$$\mathbf{F} = [\mathbf{e}']_\times [\mathbf{m}]_\times [\mathbf{e}]_\times . \tag{10}$$

A different parameterization of the fundamental matrix for planar motion is given by Vieville and Lingrand in [1995].

2.3 Calibration Constraints

The fundamental matrix $\mathbf{F}$ describes the epipolar geometry between a stereo pair. It is also known as the uncalibrated version of the essential matrix $\mathbf{E}$ described by Longuet-Higgins [Longuet-Higgins, 1981]. Both $\mathbf{F}$ and $\mathbf{E}$ are rank 2 matrices. For an arbitrary stereo pair the rank 2 constraint is the only constraint on the fundamental matrix.

When the intrinsic parameters remain constant and the relative orientation is described by planar motion, an additional constraint is imposed on the fundamental matrix. From a result due to Maybank [Maybank, 1993], the symmetric part of the essential matrix, $\mathbf{E} + \mathbf{E}^\mathbf{T}$, is rank 2 for planar motion. It is simple to show that when the intrinsic parameters remain constant this can be extended to the uncalibrated case, providing the following additional constraint on the fundamental matrix,

$$\det(\mathbf{F} + \mathbf{F}^\mathbf{T}) = 0. \tag{11}$$

This constraint reduces the number of free parameters in the fundamental matrix from 7 to 6. Note that the parameterization given by (10) enforces the above constraint and also depends upon 6 parameters, 2 for each of $\mathbf{e}$, $\mathbf{e}'$ and $\mathbf{m}$. The fewer degrees of freedom in the fundamental matrix for catadioptric stereo will lead to more robust estimates.

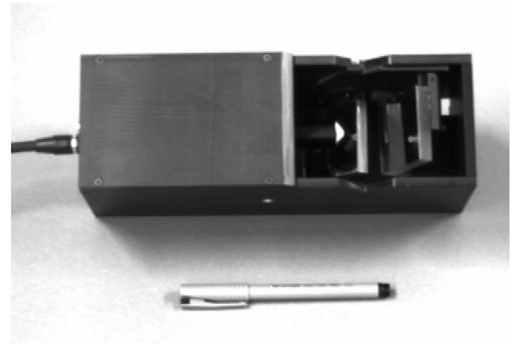

Figure 3: Catadioptric stereo system. By imaging two planar mirrors with a single camera, this compact unit outputs stereo images embedded in a single signal.

Once the epipolar geometry is found constraints can also be placed on the affine calibration. Affine calibration is achieved by identifying the homography of the plane at infinity $\mathbf{H}_\infty$ (uncalibrated rotation) [Luong and Vieville, 1996]. Given the Fundamental matrix there are still three unknown parameters needed to recover $\mathbf{H}_\infty$. To estimate these parameters it is necessary to find correspondences of points or lines at infinity. For planar motion, the horizon line of the plane of motion is the same for both images, and can be computed from the image as the line containing the two epipoles [Armstrong *et al.*, 1996]. This provides one line correspondence and thus reduces the unknown affine parameters by one. In addition, the modulus constraint described in [Pollefeys and Gool, 1997] provides a polynomial constraint on the remaining affine parameters. This constraint is derived from the observation that when the intrinsic calibration parameters are constant, $\mathbf{H}_\infty$ is conjugated with a rotation matrix. In summary, catadioptric stereo with planar mirrors introduces constraints which reduce the number of degrees of freedom in both the epipolar and the affine geometry of the stereo pair, thus leading to more stable numerical results.

3 Real-Time Implementation

Real-time stereo systems have been implemented by several researchers [Faugeras *et al.*, 1993] [Matthies, 1993]

Figure 4: Estimated epipolar geometry. The epipolar geometry was computed using the 8-point linear algorithm and then enforcing the planar motion constraint by non-linear minimization. The two bright lines indicate the estimated horizon line of the planar motion and the estimated image of the screw axis (intersection of the mirrors).

[Kanade *et al.*, 1996] [Konolige, 1997]. All of these systems use two or more cameras to acquire stereo data. Here, we describe a real-time catadioptric stereo system which uses a single camera. Figure 3 shows a picture of the catadioptric stereo system we have designed. A single Sony XC-75 b/w camera is used with two high quality Melles Griot $2''$ mirrors.

3.1 Calibration and Rectification

To achieve real-time performance it is necessary to have scanline correspondence between the stereo pair. This allows stereo matching algorithms to be implemented efficiently as described by Faugeras *et al.* [1993]. Because catadioptric stereo requires rotated mirrors (if only two mirrors are used), we must rectify the stereo pair at run-time. To compute the rectification transform we first need to estimate the fundamental matrix.

An initial estimate $\hat{\mathbf{F}}$ of the fundamental matrix is found using manual correspondences and the 8-point algorithm of [Hartley, 1995]. We then enforce the planar motion constraint (11) by performing non-linear optimization using the parameterization defined in (10). Initialization of $\mathbf{e}$, $\mathbf{e}'$ and $\mathbf{m}$ is found by extracting the epipoles and the image of the screw axis from $\hat{\mathbf{F}}$ using the method described in [Armstrong, 1996]. The error criteria minimized is the sum of squared distances to epipolar lines and the Levenberg-Marquardt algorithm is used to perform the minimization. Figure 4 shows an example of the estimated epipolar geometry. Note that it is not necessary to enforce the planar motion constraint, however the epipolar geometry for planar motion depends upon fewer degrees of freedom and thus is more resistant to noise in the correspondences.

After computing the fundamental matrix, we find a rectification transform using the method of Hartley and Gupta [Hartley and Gupta, 1993]. Once computed, this transform is used to warp each incoming image at run-time. The brightness value of each pixel in the warped image is determined by back projecting to the input image through the rectification transform and bilinearly interpolating among adjacent pixels.

3.2 Stereo Matching

The underlying assumption of all stereo matching algorithms is that the two image projections of a scene patch are similar. The degree of similarity is computed using a variety of measures such as brightness, texture, color, edge orientation, etc. Due to computational demands, most real-time systems use a measure of similarity based on image brightness. However, differences in focal settings, lens blur and gain control between the two cameras results in the two patches having different intensities. For this reason many methods such as normalized cross-correlation, Laplacian of Gaussian, and normalized sum of squared differences have been developed which attempt to compensate for camera differences. By using a single camera, catadioptric stereo avoids both the computational cost and loss of information which results from using these methods.

One of the simplest measures of similarity between two image patches is the sum of absolute differences (SAD). Because we use only a single camera, SAD is a suitable choice. SAD keeps the data size small and is easily implemented on SIMD (single instruction multiple data) processors such as those with MMX technology. Furthermore, SAD lends itself to efficient scanline correspondence algorithms.

Stereo matches are found by using a standard window based search. The search is limited to an interval of 32 pixels along the epipolar line (scanline) of a 320×240 image. By using a simple measure of similarity (SAD), scanline correspondence, and SIMD instructions we were able to achieve a throughput of approximately 20 fps on a 300Mhz Pentium II machine. An example catadioptric stereo image and computed depth map is shown in figure 5.

4 Future Directions

We have examined the geometry of stereo with two planar mirrors and shown that the epipolar geometry is restricted to the class of planar motions. In addition we have implemented a real-time stereo system using a single camera and two planar mirrors. By using methods from “uncalibrated stereo” [Hartley and Gupta, 1993] [Hartley, 1995] we have shown that catadioptric stereo can be performed with two mirrors in an arbitrary configuration.

Although single camera stereo eliminates inter-camera differences intra-camera differences still remain. In the future we intend to investigate intra-camera effects such

Figure 5: Stereo image and depth map. On the left is an image taken by a catadioptric stereo system and on the right is the depth map computed with a 7×7 correlation window.

differences across the CCD and the $\cos^4(\alpha)$ decay in image irradiance. Both of these may result in different intensities at corresponding image points. However, through calibration these effects can be measured and removed.

Other future directions include the incorporation of color and control of the aperture to improve the stereo data. Catadioptric stereo may benefit from color because only a single color camera needs to be used and therefore differences in color response curves are not a factor. Aperture control may provide additional information for stereo matching. By obtaining multiple images with different aperture settings we can increase the dynamic range of the stereo camera. Again, we need not worry about differences in aperture settings between the two virtual cameras.

In conclusion, we feel that the sensor used to acquire the stereo data is just as important as the algorithm used for matching. In this respect, catadioptric stereo offers a significant benefit by improving the quality of the stereo data at no additional computational cost.

References

[Armstrong *et al.*, 1996] M. Armstrong, A.Zisserman, and R. Hartley. Self-calibration from image triplets. In *ECCV '96*, 1996.

[Armstrong, 1996] M. Armstrong. *Self-calibration from image sequences*. PhD thesis, University of Oxford, 1996.

[Beardsley and Zisserman, 1995] P.A. Beardsley and A. Zisserman. Affine calibration of mobile vehicles. In R. Mohr and W. Chengke, editors, *Europe-China Workshop on Geometrical Modelling and Invariants for Computer Vision, Xi'an, China*, 1995.

[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 correalation-based stereo: algorithm, implementation and application. Technical Report 2013, INRIA Sophia Antipolis, 1993.

[Gluckman *et al.*, 1998] J. Gluckman, S.K. Nayar, and K.J. Thoresz. Real-time omnidirectional and panoramic stereo. In *Proceedings of the 1998 DARPA Image Understanding Workshop*, 1998.

[Goshtasby and Gruver, 1993] A. Goshtasby and W.A. Gruver. Design of a single-lens stereo camera system. *Pattern Recognition*, 26(6):923–937, 1993.

[Hartley and Gupta, 1993] R.I. Hartley and R. Gupta. Computing matched-epipolar projections. In *Proceedings of the 1993 Conference on Computer Vision and Pattern Recognition*, 1993.

[Hartley, 1995] R.I. Hartley. In defense of the 8-point algorithm. In *Proceedings of the 5th Internation Conference on Computer Vision*, pages 1064–1070, 1995.

[Hecht and Zajac, 1974] E. Hecht and A. Zajac. *Optics*. Addison-Wesley, 1974.

[Inaba *et al.*, 1993] M. Inaba, T. Hara, and H. Inoue. A stereo viewer based on a single camera with view-control mechanism. In *Proceedings of the International Conference on Robots and Systems*, July 1993.

[Kanade *et al.*, 1996] T. Kanade, A. Yoshida, K. Oda, H. Kano, and M. Tanaka. Proceedings of the 1996 conference on computer vision and pattern recognition. In *A stereo machine for video-rate dense depth mapping and its new applications*, 1996.

[Konolige, 1997] K. Konolige. Small vision systems: hardware and implementation. In *8th Int'l Symposium of Robotics Research*, Hayama, Japan, Oct. 1997.

[Longuet-Higgins, 1981] H.C. Longuet-Higgins. A computer algorithm for recunstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[Luong and Vieville, 1996] Q.T. Luong and T. Vieville. Canoncial representations for the geometries of multiple projective views. *Computer vision and image understanding*, 64(2):193–229, Sept. 1996.

[Mathieu and Devernay, 1993] H. Mathieu and F. Devernay. Systeme de miroirs pour la stereoscopie. Technical Report 0172, INRIA Sophia-Antipolis, 1993. in French.

[Matthies, 1993] L. Matthies. Stereo vision for planetary rovers: stochastic modeling to near realtime implementation. *International Journal of Computer Vision*, 8(1):71–91, 1993.

[Maybank, 1993] S.J. Maybank. *Theory of reconstruction from image motion*. Spinger-Verlag, Berlin, 1993.

[Pollefeys and Gool, 1997] M. Pollefeys and L.V. Gool. A stratified approach to metric self-calibration. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition*, 1997.

[Vieville and Lingrand, 1995] T. Vieville and D. Lingrand. Using singular displacements for uncalibrated monocular visual systemts. Technical Report 2678, INRIA Sophia-Antipolis, 1995.