



## Catadioptric Stereo Using Planar Mirrors

JOSHUA GLUCKMAN AND SHREE K. NAYAR

*Department of Computer Science, Columbia University, New York, NY 10027*

*Received July 23, 1999; Revised January 12, 2001; Accepted April 23, 2001*

**Abstract.** By using mirror reflections of a scene, stereo images can be captured with a single camera (catadioptric stereo). In addition to simplifying data acquisition single camera stereo provides both geometric and radiometric advantages over traditional two camera stereo. In this paper, we discuss the geometry and calibration of catadioptric stereo with two planar mirrors. In particular, we will show that the relative orientation of a catadioptric stereo rig is restricted to the class of planar motions thus reducing the number of external calibration parameters from 6 to 5. Next we derive the epipolar geometry for catadioptric stereo and show that it has 6 degrees of freedom rather than 7 for traditional stereo. Furthermore, we show how focal length can be recovered from a single catadioptric image solely from a set of stereo correspondences. To test the accuracy of the calibration we present a comparison to Tsai camera calibration and we measure the quality of Euclidean reconstruction. In addition, we will describe a real-time system which demonstrates the viability of stereo with mirrors as an alternative to traditional two camera stereo.

**Keywords:** stereo vision, real-time stereo, sensors, catadioptric mirrors

### 1. Introduction

Optical systems consisting of a combination of refracting (lens) and reflecting (mirror) elements are called *catadioptric* systems (Hecht and Zajac, 1974). Stereo is one area of computer vision which can benefit from such systems. By using two or more mirrored surfaces, a stereo view can be captured by a single camera (catadioptric stereo). This has the following advantages over traditional two camera stereo.

- **Identical System Parameters:** Lens, CCD and digitizer parameters such as blurring, lens distortions, focal length, spectral response, gain, offset, pixel size, etc. are identical for the stereo pair. Having identical system parameters facilitates stereo matching.
- **Ease of Calibration:** Because only a single camera and digitizer are used, there is only one set of intrinsic calibration parameters. Furthermore, we will show that the extrinsic calibration parameters are constrained by planar motion. Together these constraints

reduce the total number of calibration parameters from 16 in traditional stereo to 10 in our case.

- **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.

With these advantages in mind, we have designed and implemented a real-time catadioptric stereo system which uses only a single camera and two planar mirrors. In addition, we have analyzed the geometry and calibration of stereo with planar mirrors placed in an arbitrary configuration.

### 2. Previous Work

Previously, several researchers have demonstrated the use of both curved and planar mirrors to acquire stereo data. Curved mirrors have been primarily used to capture a wide field of view. One of the first uses of curved 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. (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.

Nayar and Baker (1997) derived the class of mirrors which produce a single view point when imaged by a camera. Later, Nene and Nayar (1998) presented several different catadioptric stereo configurations using a single camera with planar, parabolic, elliptic, and hyperbolic mirrors. A catadioptric stereo system using hyperbolic mirrors was implemented by Chaen et al. (1997). Gluckman et al. (1998) demonstrated a real-time panoramic stereo system using two coaxial catadioptric cameras (with parabolic mirrors).

The use of planar mirrors to acquire multi-view data has also been investigated. As pointed out by several researchers (Teoh and Zhang, 1984; Nishimoto and Shirai, 1987; Murray, 1995), it is possible to reconstruct a scene by imaging the scene reflection in a rotating planar mirror. However, these systems require more than one image and therefore a static scene. Mitsumoto et al. (1992) previously described a stereo method which images an object and its reflections in a set of planar mirrors. Here, the mirrors were used to obtain occlusion free images of the object. A similar method was also proposed in Zhang and Tsui (1998). Planar mirrors arranged in a pyramid can also be used to obtain omnirectional stereo data as shown by Kawanishi et al. (1998). Recently, Shashua suggested using catadioptric stereo for non-rigid stereo platforms (Shashua, 1998).

Most similar to our work, are systems which capture stereo data with planar mirrors and one camera. Goshtasby and Gruver (1993) designed a single camera stereo system using a pair of planar mirrors connected by a hinge, leaving one degree of freedom between the mirrors. Their system required the hinge to be vertically aligned and centered in the image.

In the context of active vision, Inaba et al. (1993) built a single camera stereo system using four planar mirrors. They pointed out that for active vision applications, such as high speed 3-d tracking, perfect stereo synchronization is needed. Four mirrors were used so that the vergence angle could be controlled by changing the angle between two of the mirrors and gaze could be

directed in front of the camera. A similar four mirror catadioptric system was described by Mathieu and Devernay (1995).

Although several catadioptric stereo designs have been proposed in the literature, there has been no systematic analysis of the properties (geometric and radiometric), benefits and applications of such systems. In this paper, we will discuss several calibration issues in regard to *single* camera stereo with *planar* mirrors, including the determination of relative orientation, epipolar geometry, and focal length. These results together provide a theoretical foundation for planar catadioptric stereo. In addition, we will describe a real-time catadioptric stereo system which demonstrates the viability of stereo with mirrors as an alternative to traditional two camera stereo.

### 3. Geometry and Calibration

The results in this section pertain to the geometry and calibration of catadioptric stereo systems that use a single perspective camera and two planar mirrors placed in an arbitrary configuration. Figure 1 depicts the image formation of such a system. 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  $\mathbf{c}$  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 focal length of each virtual camera is equal to  $f$ , the focal length of the real camera. 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.

We will derive both the relative orientation and the epipolar geometry of catadioptric stereo with two mirrors. Then, we will discuss self-calibration constraints which can be used to recover the focal length solely from a set of image correspondences. Finally, we examine the field of view of two mirror stereo.

#### 3.1. Relative Orientation

In traditional stereo, the two cameras can be placed in any configuration, and therefore the relative orientation between the cameras is described by 6 parameters (3 for rotation and 3 for translation). For catadioptric stereo, the relative orientation between the two virtual

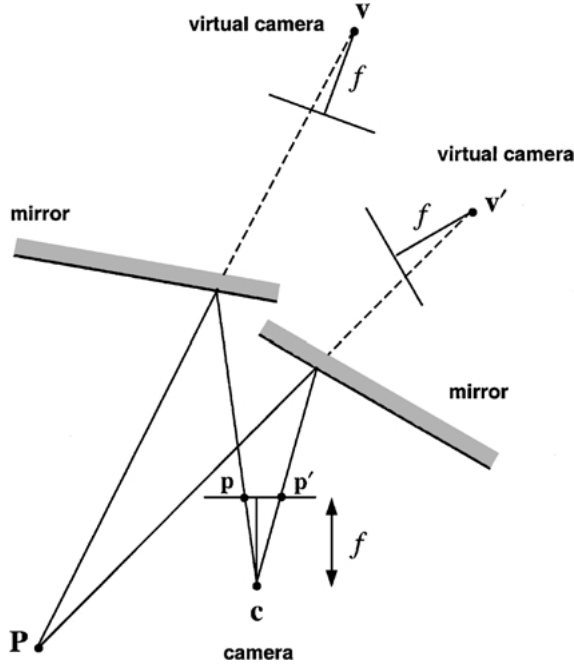


Figure 1. Stereo image formation with a single camera and two planar mirrors. A scene point  $P$  reflected by two mirrors is imaged as if seen from two different viewpoints  $v$  and  $v'$ . With knowledge of the relative orientation between  $v$  and  $v'$  and the image pair  $p$  and  $p'$ , we can triangulate and recover  $P$ .

cameras, regardless of the configuration of the mirrors, 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 rigid transformation  $\mathbf{D}$  between the two reflected viewpoints  $v$  and  $v'$ . Each virtual viewpoint is related to the camera center  $c$  as,

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

and

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

where  $\mathbf{D}_1$  and  $\mathbf{D}_2$  are reflection transformations defined in (??). Then, the relative orientation  $\mathbf{D}$  is simply,

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

Representing the two mirrors as planes with normals  $\mathbf{n}_1$  and  $\mathbf{n}_2$  and distances  $d_1$  and  $d_2$  measured from the

camera center  $c$ , the reflection transformations for the two mirrors are given by

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

and

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

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

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

where

$$\mathbf{R} = \mathbf{I} + 4(\mathbf{n}_1 \cdot \mathbf{n}_2)\mathbf{n}_1\mathbf{n}_2^T - 2\mathbf{n}_1\mathbf{n}_1^T - 2\mathbf{n}_2\mathbf{n}_2^T, \quad (1)$$

and

$$\mathbf{t} = 2d_1\mathbf{n}_1 - (4d_1(\mathbf{n}_1 \cdot \mathbf{n}_2) + 2d_2)\mathbf{n}_2. \quad (2)$$

It can be shown that  $(\mathbf{n}_1 \times \mathbf{n}_2) = \mathbf{R}(\mathbf{n}_1 \times \mathbf{n}_2)$ , therefore the rotation matrix  $\mathbf{R}$  has a rotational axis of  $\mathbf{n}_1 \times \mathbf{n}_2$ . From (2) 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. This type of motion is termed *planar motion* (Armstrong et al., 1996).

Figure 2 shows how the relative orientation between the two virtual cameras is constrained by planar motion. This constraint arises from the fact that the virtual cameras are related by a pure rotation about the axis  $\mathbf{S}$  (called a *screw axis*) which is the intersection of the planes containing the two mirrors. A pure rotation about  $\mathbf{S}$  constrains the motion between the virtual cameras to lie in a plane perpendicular to  $\mathbf{S}$ . We will be referring to the plane which contains the camera centers (both real and virtual) as the *plane of motion*. This plane is orthogonal to the screw axis and its intersection with the image plane is termed the *horizon line* of the plane of motion.

As we have seen, single camera stereo with two planar mirrors constrains the external calibration parameters to planar motion, reducing the number of parameters from 6 for conventional two camera stereo to 5

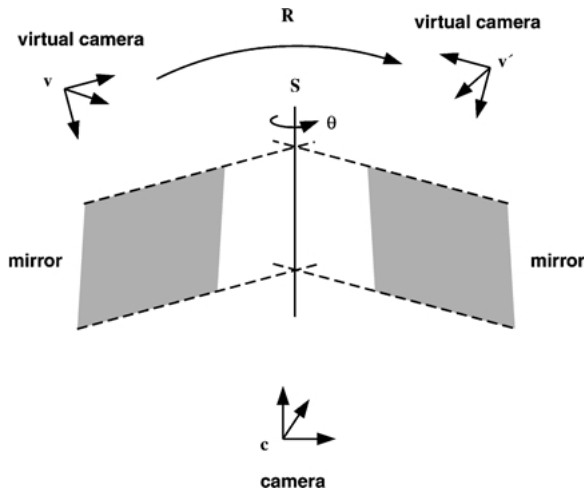


Figure 2. The relative orientation  $\mathbf{R}$  between the two virtual camera coordinate systems is described by a pure rotation  $\theta$  about the axis  $\mathbf{S}$  (called a screw axis) which is the intersection of the planes containing the two mirrors. A pure rotation about this axis constrains the translation between the virtual cameras to lie in a plane perpendicular to the screw axis, and is called planar motion. This planar motion constraint holds true for any configurations of the two mirrors.

for catadioptric stereo. Because only a single camera is used, the internal parameters (focal length, pixel size, image center, skew) are exactly the same for the two stereo views, reducing the number parameters from 10 to 5. Together, these constraints on both the external and internal parameters place restrictions on the epipolar geometry.

### 3.2. Epipolar Geometry

Epipolar geometry is a description of the geometric relationship between a pair of stereo images. It is represented by the fundamental matrix  $\mathbf{F}$  and is the minimal information necessary to determine the epipolar lines (Faugeras, 1992). For a pair of image correspondences  $\mathbf{p}$  and  $\mathbf{p}'$ ,  $\mathbf{F}$  introduces the following well-known epipolar constraint:

$$\mathbf{p}'^T \mathbf{F} \mathbf{p} = 0. \quad (3)$$

In general,  $\mathbf{F}$  is dependent on the 16 extrinsic (relative orientation) and intrinsic calibration parameters. However, for an arbitrary stereo pair  $\mathbf{F}$  only has 7 free parameters (Xu and Zhang, 1996). By constraining the extrinsic and intrinsic parameters, we will show catadioptric stereo reduces the number of free parameters in  $\mathbf{F}$  to 6.

$\mathbf{F}$  is also known as the uncalibrated version of the essential matrix  $\mathbf{E}$  described by Longuet-Higgins (1981), because

$$\mathbf{F} = \mathbf{A}'^{-T} \mathbf{E} \mathbf{A}^{-1}, \quad (4)$$

where  $\mathbf{A}'$  and  $\mathbf{A}$  are matrices representing the internal calibration parameters of the stereo cameras. 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.

From a result due to Maybank (1993), it is known that one of the eigenvalues of the symmetric part of the essential matrix,  $\mathbf{E} + \mathbf{E}^T$  is

$$\mathbf{t} \cdot \mathbf{r} \sin(\theta) \quad (5)$$

where  $\mathbf{t}$  is the direction of translation,  $\mathbf{r}$  is the axis of rotation and  $\theta$  is the angle between them. When  $\mathbf{t}$  is orthogonal to  $\mathbf{r}$ , as in planar motion, the eigenvalue is zero and thus the matrix  $\mathbf{E} + \mathbf{E}^T$  is rank deficient. When the intrinsic parameters for the two views are identical ( $\mathbf{A}' = \mathbf{A}$ ), which is true for catadioptric stereo, it is simple to extend this to the symmetric part of the fundamental matrix, providing the following additional constraint on the fundamental matrix,

$$\det(\mathbf{F} + \mathbf{F}^T) = 0. \quad (6)$$

This constraint reduces the number of free parameters in the fundamental matrix from 7 to 6 and has been used by Armstrong et al. (1996) and Vieville and Lingrand (1995) to help constrain the self-calibration of a camera mounted on a mobile robot, where ground motion can be modeled by planar motion.

When estimating the fundamental matrix from image correspondences it is useful to have a parameterization of  $\mathbf{F}$  which implicitly enforces (6). We can derive such a parameterization by considering the image projection of the screw axis.

The location of the screw axis (see Fig. 2) is the same with respect to the coordinate systems of the two virtual cameras, therefore  $\mathbf{m}$  its location in the image is identical for both the left and right stereo views. This implies that corresponding epipolar lines must intersect on  $\mathbf{m}$ . The resulting epipolar geometry is depicted in Fig. 3. As shown in this figure, the epipolar line of a point  $\mathbf{p}$  is the line containing epipole  $\mathbf{e}'$  and the intersection of  $\mathbf{m}$  with the line through epipole  $\mathbf{e}$  and point  $\mathbf{p}$ .

Using homogenous coordinates, a line containing two points is represented by the cross product of the two

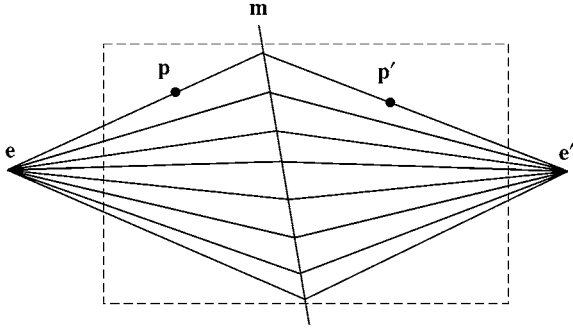


Figure 3. The epipolar geometry due to planar motion. When motion is constrained to lie in a plane, all corresponding epipolar lines must intersect at  $\mathbf{m}$  the image projection of the screw axis. Therefore, the two epipoles  $\mathbf{e}$  and  $\mathbf{e}'$  and the line  $\mathbf{m}$  completely determine the epipolar geometry.

points, and the intersection of two lines is represented by the cross product of the two vectors which represent the lines. We can therefore represent the line through  $\mathbf{e}$  and  $\mathbf{p}$  as

$$(\mathbf{e} \times \mathbf{p})$$

and the intersection of this line with the line  $\mathbf{m}$  as

$$(\mathbf{m} \times (\mathbf{e} \times \mathbf{p})).$$

The epipolar line containing this point and  $\mathbf{e}'$  is

$$(\mathbf{e}' \times (\mathbf{m} \times (\mathbf{e} \times \mathbf{p}))).$$

We can express the epipolar constraint between this line and the point  $\mathbf{p}'$  as

$$\mathbf{p}' \cdot (\mathbf{e}' \times (\mathbf{m} \times (\mathbf{e} \times \mathbf{p}))) = 0.$$

Using the relation  $[\mathbf{v}]_{\times} \mathbf{x} = \mathbf{v} \times \mathbf{x}$  for all vectors  $\mathbf{x}$  this equation is rewritten as

$$\mathbf{p}'^T [\mathbf{e}']_{\times} [\mathbf{m}]_{\times} [\mathbf{e}]_{\times} \mathbf{p} = 0.$$

From the above equation and (3), the fundamental matrix  $\mathbf{F}$  has the form

$$\mathbf{F} = [\mathbf{e}']_{\times} [\mathbf{m}]_{\times} [\mathbf{e}]_{\times}. \quad (7)$$

Each of  $\mathbf{e}$ ,  $\mathbf{e}'$  and  $\mathbf{m}$  is only defined up to a scale factor and therefore described by two parameters, giving a total of 6 parameters for  $\mathbf{F}$ .

With the help of a symbolic algebra package, we have confirmed that the parameterization given in (7) does indeed enforce the planar motion constraint (6). Other parameterizations of the fundamental matrix for planar motion are also possible, see for instance Vieville and Lingrand (1995).

Using (7) and a set of image correspondences,  $\mathbf{F}$  can be determined by searching the parameter space of  $\mathbf{e}$ ,  $\mathbf{e}'$  and  $\mathbf{m}$  while minimizing a suitable cost function such as the sum of distances of corresponding points from their epipolar lines. This process requires non-linear minimization and thus initial estimates of  $\mathbf{e}$ ,  $\mathbf{e}'$  and  $\mathbf{m}$  are needed.

Initial estimates of  $\mathbf{e}$ ,  $\mathbf{e}'$  and  $\mathbf{m}$  can be extracted from an estimate of  $\mathbf{F}$  obtained by the linear 8-point algorithm (Hartley, 1995).  $\mathbf{e}$  and  $\mathbf{e}'$  can be extracted from the left and right null space of  $\mathbf{F}$ . Using the following method,  $\mathbf{m}$  can be obtained from the eigenvectors of the symmetric part of fundamental matrix  $\mathbf{F}_s = \mathbf{F} + \mathbf{F}^T$  (Armstrong, 1996). Letting  $\lambda_1$ ,  $\lambda_2$  and  $\mathbf{n}_1$ ,  $\mathbf{n}_2$  be the positive and negative eigenvalues and eigenvectors of  $\mathbf{F}_s$ , we have either

$$\mathbf{m} = \sqrt{\lambda_1} \mathbf{n}_1 + \sqrt{-\lambda_2} \mathbf{n}_2 \quad (8)$$

or

$$\mathbf{m} = \sqrt{\lambda_1} \mathbf{n}_1 - \sqrt{-\lambda_2} \mathbf{n}_2. \quad (9)$$

The above ambiguity in  $\mathbf{m}$  can be resolved by noting that one of these expressions is equivalent to  $\mathbf{e} \times \mathbf{e}'$  and  $\mathbf{m}$  is the other one.

As we have shown, catadioptric stereo with planar mirrors introduces an additional constraint on the fundamental matrix which reduces the number of parameters to estimate from 7 to 6. Next, we will discuss recovering the focal length from a single catadioptric stereo image.

### 3.3. Recovering the Focal Length

With knowledge of the fundamental matrix, the scene geometry can be reconstructed up to an unknown projective transform (Faugeras, 1992). To obtain a Euclidean reconstruction from a stereo pair, it is necessary to determine the internal camera parameters. With video cameras, it is often the case that the aspect ratio is known, the skew is zero, and the image center is roughly the center of the image; therefore, Euclidean reconstruction amounts to determining the focal lengths of

the cameras. Through the Kruppa equations (Zeller and Faugeras, 1996), the fundamental matrix places two quadratic constraints on the internal calibration parameters. As demonstrated by Hartley (1992), these two constraints are sufficient to solve for the focal lengths when the other internal parameters are known.

For catadioptric stereo, we have only one unknown focal length  $f$  and we can solve for  $f$  from the Kruppa equations

$$\mathbf{F}\omega\mathbf{F}^T = x[\mathbf{e}' ]_{\times}\omega[\mathbf{e}' ]_{\times}, \quad (10)$$

where

$$\omega = \begin{pmatrix} f^2 & 0 & 0 \\ 0 & f^2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and  $x$  is an unknown scale factor ( $\mathbf{F}$  and  $\mathbf{e}'$  are projective quantities and thus only known up to a scale factor).

Though  $f$  can be determined in this manner it has been shown that the Kruppa equations are very unstable in practice (Zeller and Faugeras, 1996), thus we would like to explore additional constraints on the focal length that arise from the planar motion. It turns out there are two such constraints. The first results from the fact that the plane of motion is always perpendicular to the plane that contains the screw axis and the camera center. The plane of motion projects to the horizon line ( $\mathbf{e} \times \mathbf{e}'$ ) and the plane containing the screw axis and the camera center projects to  $\mathbf{m}$  (the image of the screw axis). The 3-D angle between the visual planes of two image lines  $\mathbf{x}$  and  $\mathbf{y}$  is given by Triggs (1997),

$$\frac{(\mathbf{x}^T \omega \mathbf{y})}{\sqrt{(\mathbf{x}^T \omega \mathbf{x})(\mathbf{y}^T \omega \mathbf{y})}} = \cos \theta. \quad (11)$$

Letting  $\mathbf{l} = (\mathbf{e} \times \mathbf{e}')$  we have the following constraint

$$(\mathbf{l}^T \omega \mathbf{m}) = 0, \quad (12)$$

because the plane defined by  $\mathbf{l}$  and the plane defined by  $\mathbf{m}$  must be perpendicular.

A second constraint can be derived from the image points  $\mathbf{e}$ ,  $\mathbf{e}'$ , and the point  $\mathbf{m}' = \mathbf{l} \times \mathbf{m}$ , which is the intersection of the image of the screw axis  $\mathbf{m}$  and the horizon line  $\mathbf{l}$ . From Fig. 4, we can see that the angle formed between the image rays through  $\mathbf{e}$  and  $\mathbf{m}'$  is equal to the angle formed by  $\mathbf{e}'$  and  $\mathbf{m}'$ . Using a relationship similar to (11) but for image rays (Triggs,

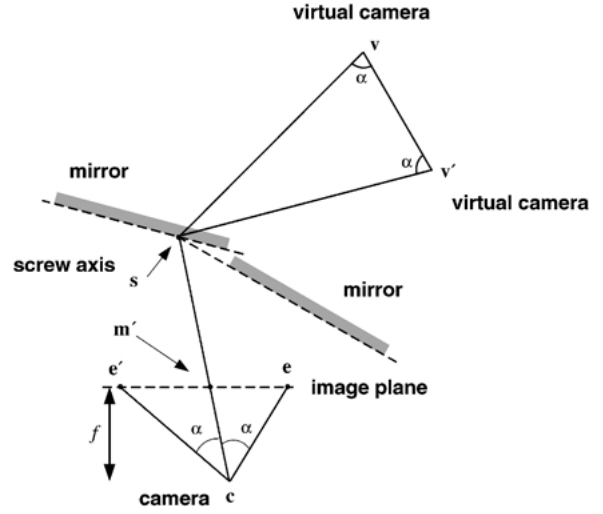


Figure 4. In the plane of motion (the plane defined by  $\mathbf{c}$ ,  $\mathbf{v}$ ,  $\mathbf{v}'$ ) the two virtual camera centers will be the same distance from  $\mathbf{s}$  the screw axis. As a result, the angles formed by  $\mathbf{v}$  and  $\mathbf{v}'$  and  $\mathbf{s}$  are equivalent. The image projection of  $\mathbf{v}$  and  $\mathbf{v}'$  are the epipoles and  $\mathbf{s}$  projects to  $\mathbf{m}'$ , therefore the angles formed by  $\mathbf{e}$ ,  $\mathbf{e}'$  and  $\mathbf{m}'$  are also equivalent. This constraint can be used to recover  $f$  the focal length provided  $\mathbf{m}'$  is not the image center.

1997), we can express this as

$$\frac{(\mathbf{e}^T \omega^{-1} \mathbf{m}')}{\sqrt{(\mathbf{e}^T \omega^{-1} \mathbf{e})(\mathbf{m}'^T \omega^{-1} \mathbf{m}')}} = \frac{(\mathbf{e}'^T \omega^{-1} \mathbf{m}')}{\sqrt{(\mathbf{e}'^T \omega^{-1} \mathbf{e}')(\mathbf{m}'^T \omega^{-1} \mathbf{m}')}} \quad (13)$$

When using these equations to recover the focal length, care must be taken to avoid degenerate configurations. In particular, when  $\mathbf{m}$  passes through the image center, (13) will not lead to a solution for  $f$  (see Fig. 5). We can ensure  $\mathbf{m}$  does not pass through the image center by displacing the mirrors as in Fig. 4. Equation (12) can not be used when  $\mathbf{m}$  and  $\mathbf{l}$  are perpendicular. Avoiding this configuration is more difficult, it requires displacing the mirrors and tilting the camera upward or downward with respect to the mirrors.

### 3.4. Field of View

Because the two virtual cameras must share the field of view the stereo system is limited to half of the field of view of the camera. Depending on the angle between the mirrors the shared field of view of the stereo system may be further limited. As shown in Fig. 6 the overlapping field of view of the stereo system is  $2\alpha$  where  $\alpha$

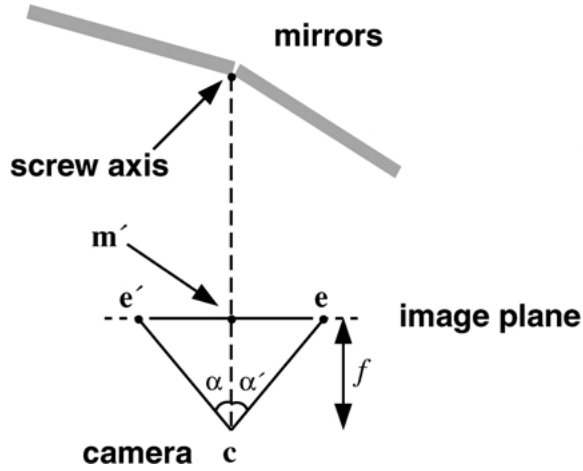


Figure 5. A degenerate configuration for which the focal length can not be recovered. When the intersection of the two mirrors projects to the image center, any focal length  $f$  satisfies the constraint that the angles  $\alpha$  and  $\alpha'$  are equal.

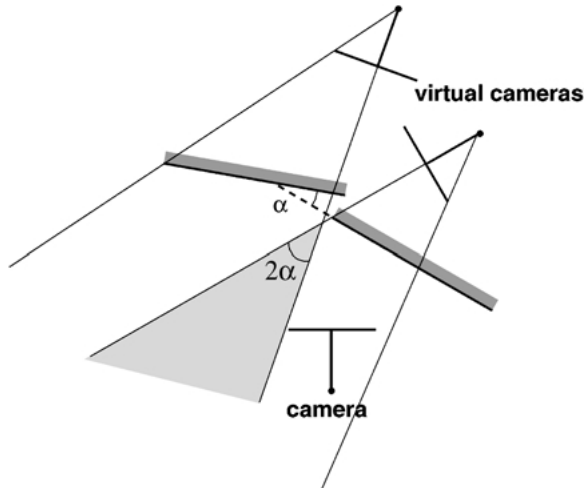


Figure 6. When two mirrors are used the common field of view of the virtual cameras is  $2\alpha$  where  $\alpha$  is the angle between the two mirrors.

is the angle between the two mirrors. In order for the shared field of view to be unbounded the camera must have a field of view of at least  $2\alpha$ .

#### 4. Experiments: Recovering the Focal Length

As shown in Section 3.3, the focal length and relative orientation of a catadioptric stereo system can be estimated from a set of correspondences taken from

a single image. To test the accuracy of the proposed method we performed both real and simulated experiments.

In the first experiment we compare the estimated focal length obtained from the angle constraint (13) to the focal length obtained from Tsai calibration (Tsai, 1986).<sup>1</sup> Note that Tsai calibration uses a set of known 3-D points and their image correspondences while the angle constraint only requires a set of image correspondences. For the second experiment, we reconstruct rectangles in 3-D and measure the angles of the four corners. Any errors in the calibration will manifest themselves as deviations from 90 degrees. Although the constraint (12) can also be used to recover the focal length, we found this constraint not as practical because the camera must be pointed up or down at an oblique angle in order to avoid degenerate configurations. In the simulations we used randomly generated image correspondences to examine the behavior of self calibration in response to varying amounts of noise, rotation, field of view and location of the screw axis. We also performed two sets of real experiments.

##### 4.1. Comparison to Tsai Calibration

We took a series of 10 catadioptric images using a Sony XC-75 camera with a Computar 4mm Pinhole lens (no radial distortions are present). For each image the mirrors were placed in a configuration similar to Fig. 4 in order to avoid  $\mathbf{m}'$  passing through the image center. Throughout the sequence we varied the angle between the mirrors and used several different scenes.

For each catadioptric image we find an initial estimate  $\hat{\mathbf{F}}$  of the fundamental matrix and a set of correspondences  $\{(\mathbf{p}_i, \mathbf{p}'_i)\}$  using the robust method of [36], which is publicly available.<sup>2</sup> From  $\hat{\mathbf{F}}$  initial estimates of  $\mathbf{e}$  and  $\mathbf{e}'$  are obtained from the left and right null space and  $\mathbf{m}$  is found using Eqs. (8) and (9). We then enforce the planar motion constraint (6) by performing non-linear optimization on the parameters  $\mathbf{e}$ ,  $\mathbf{e}'$  and  $\mathbf{m}$  using the parameterization defined in (7). The error criteria minimized is the sum of squared distances of the corresponding points to their epipolar lines. Defining  $d(\mathbf{p}'_i, \mathbf{F}\mathbf{p}_i)$  to be the distance of point  $\mathbf{p}'_i$  to the epipolar line  $\mathbf{F}\mathbf{p}_i$ , we seek the  $\mathbf{e}$ ,  $\mathbf{e}'$  and  $\mathbf{m}$  that minimize

$$\xi = \sum_i d^2(\mathbf{p}'_i, \mathbf{F}\mathbf{p}_i) + d^2(\mathbf{p}_i, \mathbf{F}^T \mathbf{p}'_i), \quad (14)$$



Figure 7. (a) A set of image correspondences used to compute the epipolar geometry. (b) The recovered epipolar geometry. The vertical thick line is  $\mathbf{m}$  the estimated image of the screw axis, where the corresponding epipolar lines meet. The horizontal thick line is the line connecting the two epipoles, the horizon line of the planar motion. The intersection of the two lines is  $\mathbf{m}'$  and is used to estimate the focal length.

where

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

We solve this non-linear minimization problem by using the Levenberg-Marquardt algorithm (Press et al., 1992). Note that each of  $\mathbf{e}$ ,  $\mathbf{e}'$  and  $\mathbf{m}$  are only defined up to a scale factor and therefore need to be parameterized by two values. This can either be done with spherical coordinates or by setting a component to one for each vector. After minimization, Eq. (13) and the estimates of  $\mathbf{e}$ ,  $\mathbf{e}'$  and  $\mathbf{m}' = \mathbf{e} \times \mathbf{e}' \times \mathbf{m}$  are used to obtain an estimate of the focal length. Figure 7 shows a typical scene with a set of correspondences and the recovered epipolar geometry.

As a comparison we use the following well known method to compute the focal length of a camera. An image is taken of a calibration box with known 3-D points (see Fig. 8). The image locations of these points are computed to sub-pixel accuracy by finding the centroid of the projected circles. Using the image locations and the corresponding known 3-D points the focal length can be recovered using Tsai's method (Tsai, 1986).

The following table shows the estimated focal lengths (in pixels) for each of the 10 catadioptric images as compared to the focal length obtained from Tsai calibration.

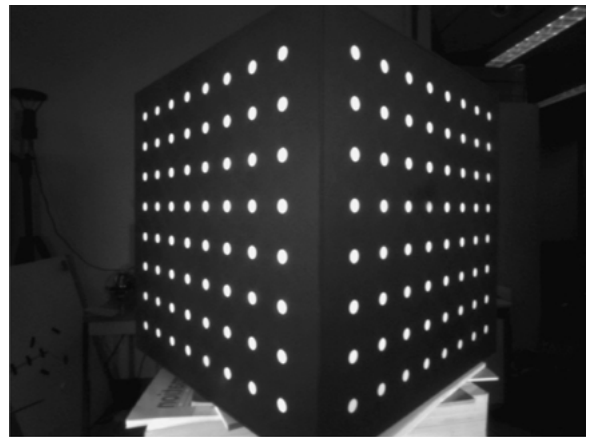


Figure 8. Tsai calibration. By imaging a set of known 3-D points we recovered the focal length through Tsai calibration and compared this to calibration using the geometric constraints of catadioptric stereo.

Catadioptric images					Tsai
440	437	439	440	437	407
433	439	420	428	423	

The discrepancy of the focal lengths for the catadioptric images as compared to Tsai is possibly due to several factors: inaccurate localization of the epipoles, bias from assuming the location of the center of projection is the image center, and/or inaccuracies in Tsai calibration.



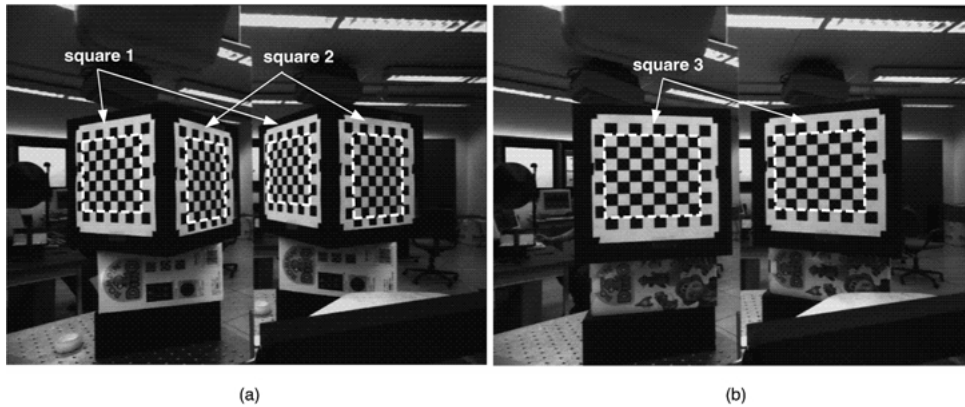


Figure 9. From corresponding points we reconstruct these squares and measure how close the angles of the four corners are to 90 degrees. (a) In this image the two squares are at an angle. (b) The second image is of a frontal square.

#### 4.2. Euclidean Reconstruction

In the second set of experiments, we determine the quality of calibration by measuring the angles of reconstructed squares. If the estimated focal length is inaccurate then the 3-D reconstruction will have projective distortion causing angles to deviate from 90 degrees.

First we estimate the fundamental matrix and the focal length of the catadioptric stereo system using the methods just described in the previous section. Using the focal length, the fundamental matrix and the method of Longuet-Higgins (1981) we are able to determine the rotation and translation up to a scale factor between the two virtual cameras.

Then, as shown in Fig. 9 we took images of squares in several different positions. Using 4 manually selected stereo correspondences for each square we reconstructed the square in 3-D and measured the angles at each corner (which should be 90 degrees). The reconstruction was performed by triangulation, using the estimated focal length and extrinsic parameters. The estimated 3-D point was taken to be the midpoint of the shortest distance between the two back projected rays. Although the direction of translation is only known up to a scale factor this does not affect the measured angles.

For the first image a total of 8 angles were measured from the two squares and 4 angles were measured from the third square in the second image (see Fig. 9). The following table contains the computed angles. For each square the angles are measured clockwise from the upper left corner.

Square 1	89.8	89.2	89.3	91.7
Square 2	91.1	91.4	87.8	89.7
Square 3	90.2	90.4	90.6	88.9

For the 12 angles measured the mean was  $90^\circ$  with standard deviation  $1.08^\circ$ . Given that we only have pixel resolution for the correspondences this is about as good as we can expect. Although the first experiment revealed that self calibrating the focal length from the catadioptric geometry is not as accurate as Tsai calibration, it is suitable for quality Euclidean reconstructions.

#### 4.3. Simulations

When self calibrating, there are many factors that may effect the estimation of the focal length. In this section we examine the effect of noise, rotation between the virtual cameras, field of view and location of the image of the screw axis via simulations. For each simulation we use an image of size  $640 \times 480$  pixels and 100 image correspondences. The epipolar geometry is determined from the location of the epipoles and the image of the screw axis. In this simulation, the optical axis of the camera is restricted to lie in the same plane as the mirror normals. Therefore, the image of the screw axis is vertical and the epipoles lie on a horizontal line through the center of the image. The location of the screw axis in the image is described by the parameter  $c$ , which is the distance from the center of the image. The location of one epipole is determined from the angle of rotation  $\theta$  which is twice the angle between the mirrors.

The other epipole is determined from the focal length  $f$ . Once the epipolar geometry is determined, image correspondences are generated by randomly selecting a pixel in the left half of the image and then randomly selecting a corresponding pixel along the epipolar line in the right half of the image. If the pair of image correspondences projects to a scene point that does not lie in front of the cameras, the pair is discarded and another pair is generated. Once the image correspondences are generated, Gaussian noise with standard deviation  $\eta$  is added to their locations in the image.

For each simulation an estimate  $\hat{f}$  of the focal length is obtained from the randomly generated image correspondences using the method described in Section 12. Each simulation is performed 100 times and the mean square error (MSE) ( $\frac{\sum(\hat{f}-f)^2}{100}$ ) is computed. Table 1 shows the results of the simulations while varying the location of the screw axis  $c$ , the amount of noise  $\eta$ , the amount of rotation  $\theta$  and the field of view as given by the focal length  $f$ . As expected, Table 1(a) shows how the MSE increases as the image of the screw axis approaches the center of the image  $c = 0$ . Table 1(b) shows the response of the MSE to increasing amounts of noise. As shown in (c) the MSE appears to not be

*Table 1.* The table shows the results from simulations of estimating the focal length from a set of randomly generated image correspondences. The effect of the following four parameters is examined: the location of the screw axis  $c$ , the amount of noise  $\eta$ , rotation between the virtual cameras  $\theta$ , and field of view as determined by the focal length  $f$ . For each simulation one parameter is varied while the other three remain constant and the mean square error of the estimated focal length is shown. The value for  $c$  is the distance in pixels from the center of the image.  $\eta$  is the standard deviation of the noise in pixels.  $\theta$  is the angle of rotation between the virtual cameras and is twice the angle between the two mirrors.

---

(a) Constant parameter values: $f = 457$ , $\eta = 0.4$ , $\theta = 10^\circ$										
Screw Axis ( $c$ )	300	270	240	210	180	150	120	90	60	30
MSE of $\hat{f}$	1.5	1.8	0.9	1.4	2.0	2.3	3.2	5.7	15.5	130.6
(b) Constant parameter values: $f = 457$ , $c = 270$ , $\theta = 10^\circ$										
Noise ( $\eta$ )	0.0	0.4	0.8	1.2	1.6					
MSE of $\hat{f}$	0.0	1.8	5.3	13.4	22.0					
(c) Constant parameter values: $f = 457$ , $c = 270$ , $\eta = 0.4$										
Rotation ( $\theta$ )	2°	6°	10°	14°	18°					
MSE of $\hat{f}$	1.5	1.6	1.4	1.1	1.3					
(d) Constant parameter values: $c = 270$ , $\eta = 0.4$ , $\theta = 10^\circ$										
Focal length ( $f$ )	300	500	700	900	1100					
MSE of $\hat{f}$	1.8	1.6	8.8	35.1	99.2					

---

affected by the angle between the mirrors. For large focal lengths and thus small fields of view the MSE increases as shown in (d).

## 5. Real-Time Implementation

Real-time stereo systems have been implemented by several researchers (Faugeras et al., 1993; Matthies, 1993; 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 and only a PC to compute depth-maps in real-time. Catadioptric stereo is well suited for real-time implementation on a PC because only a single camera and digitizer is needed. Use of a single camera obviates the need for synchronization hardware and software.

### 5.1. Sensor Design

Figure 10 shows a picture of two catadioptric stereo sensors we have designed. Both designs provide adjustments which allow the rotation and baseline between the two virtual cameras to be controlled. The schematic in Fig. 11 depicts the catadioptric stereo sensor adjustments. When the physical camera is moved away from the mirrors the two virtual cameras move along lines normal to the two mirrors, effectively increasing the stereo baseline while holding the rotation between the cameras constant. In addition, the angle between the mirrors can be adjusted. Increasing this angle results in a larger rotation and baseline between the virtual cameras.

### 5.2. Calibration and Rectification

To achieve real-time performance it is necessary to have scanline correspondence between stereo views. 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) and hence rotated views, we must rectify the stereo pair at run-time.

To compute the rectification transform we first need to estimate the fundamental matrix. An estimate of the fundamental matrix is found using the method previously described. After computing the fundamental matrix, we find rectification transforms for the left and right images, using a method based on that of Hartley

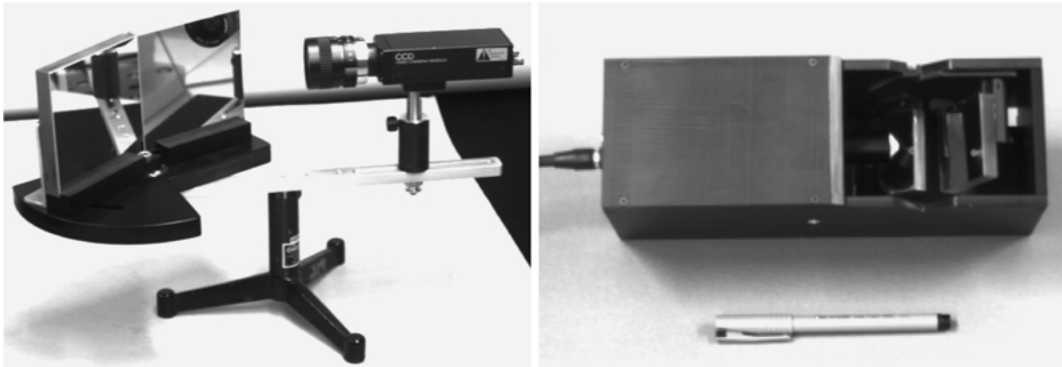


Figure 10. Catadioptric stereo sensors. (left) A single Sony XC-77 b/w camera and a 12.5 mm Computar lens is used with two high quality front silvered Melles Griot 5" mirrors. The distance between the camera and mirrors can be altered, which changes the baseline of the stereo system. The angle between the mirrors can also be adjusted to control vergence and rotation between the stereo views. (right) This compact unit uses a single Sony XC-75 b/w camera and a 4 mm Computar pinhole lens with 2" Melles Griot mirrors.

and Gupta (1993). Once computed, the rectification transforms are 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.

### 5.3. Stereo Matching

The underlying assumption of all stereo matching algorithms is that the two image projections of a small scene patch are similar. The degree of similarity is computed using a variety of measures such as bright-

ness, texture, color, edge orientation, etc. To minimize computations, 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 employed to compensate for camera differences (Faugeras et al., 1993; Matthies, 1993; Kanade et al., 1996; Konolige, 1997). By using a single camera, catadioptric stereo avoids both the computational cost and loss of information which results from using these methods.

As Fig. 12 shows, normalized cross-correlation and the Laplacian of Gaussian can degrade the performance of stereo matching due to loss of information and finite arithmetic. By removing differences in offset and gain, normalized cross-correlation and the Laplacian of Gaussian also remove absolute intensity information which is useful for matching when only a single camera is used.

Each depth map in Fig. 12 was computed using a different measure of similarity. The first measure was sum of absolute difference (SAD), the second measure was normalized cross-correlation (NCORR) and the third was SAD after application of a Laplacian of Gaussian (LOG) operator.<sup>3</sup> Due to the loss of information from the LOG operator, the third measure performed the worst. NCORR and SAD performed similar for large window sizes, greater than  $8 \times 8$ . However, for small windows the results from SAD were better than NCORR. This is in contrast to two camera stereo

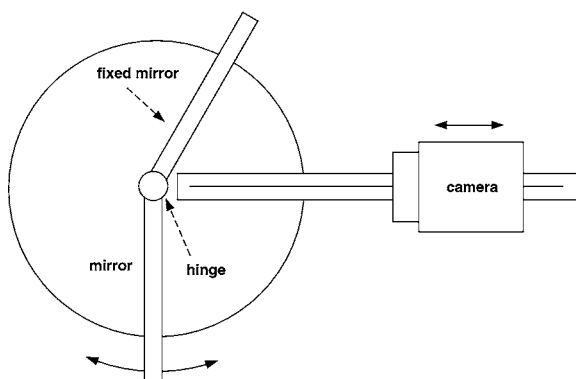
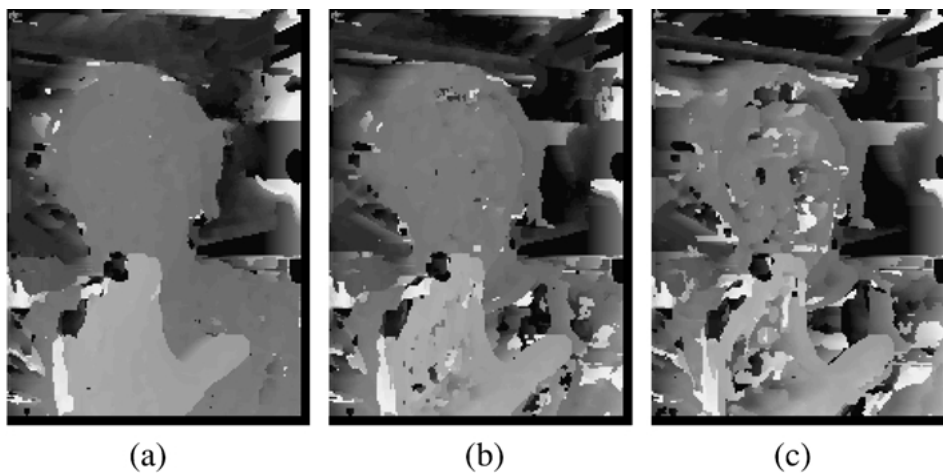


Figure 11. Schematic of catadioptric stereo sensor adjustments. By adjusting the angle of the two mirrors the rotation and baseline between the virtual cameras can be controlled. The baseline can also be altered by adjusting the distance of the physical camera to the mirrors.



*Figure 12.* Comparison of three commonly used measures of similarity on an image taken by a catadioptric stereo sensor. (a) Depth map computed using sum of absolute differences. (b) Depth map computed using normalized cross-correlation. (c) Depth map computed using sum of absolute differences after a Laplacian of Gaussian operator was applied. For all three measures a  $5 \times 5$  window was used and no thresholds were applied.

where NCORR tends to out perform SAD (Faugeras et al., 1993). From a computational standpoint, SAD is far more desirable than NCORR, in fact an approximation to NCORR was needed to achieve real-time performance in Faugeras et al. (1993).

In our implementation, we chose to use SAD as a measure of similarity for stereo matching. As we have

seen, for single camera stereo SAD has several advantages over other measures: no loss of information, small window sizes can be used and a fast implementation is possible. In addition, SAD keeps the data size small, as opposed to SSD, and is easily implemented on SIMD (single instruction multiple data) processors such as those with MMX technology. Furthermore,



*Figure 13.* 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 the SAD measure and a  $7 \times 7$  window.

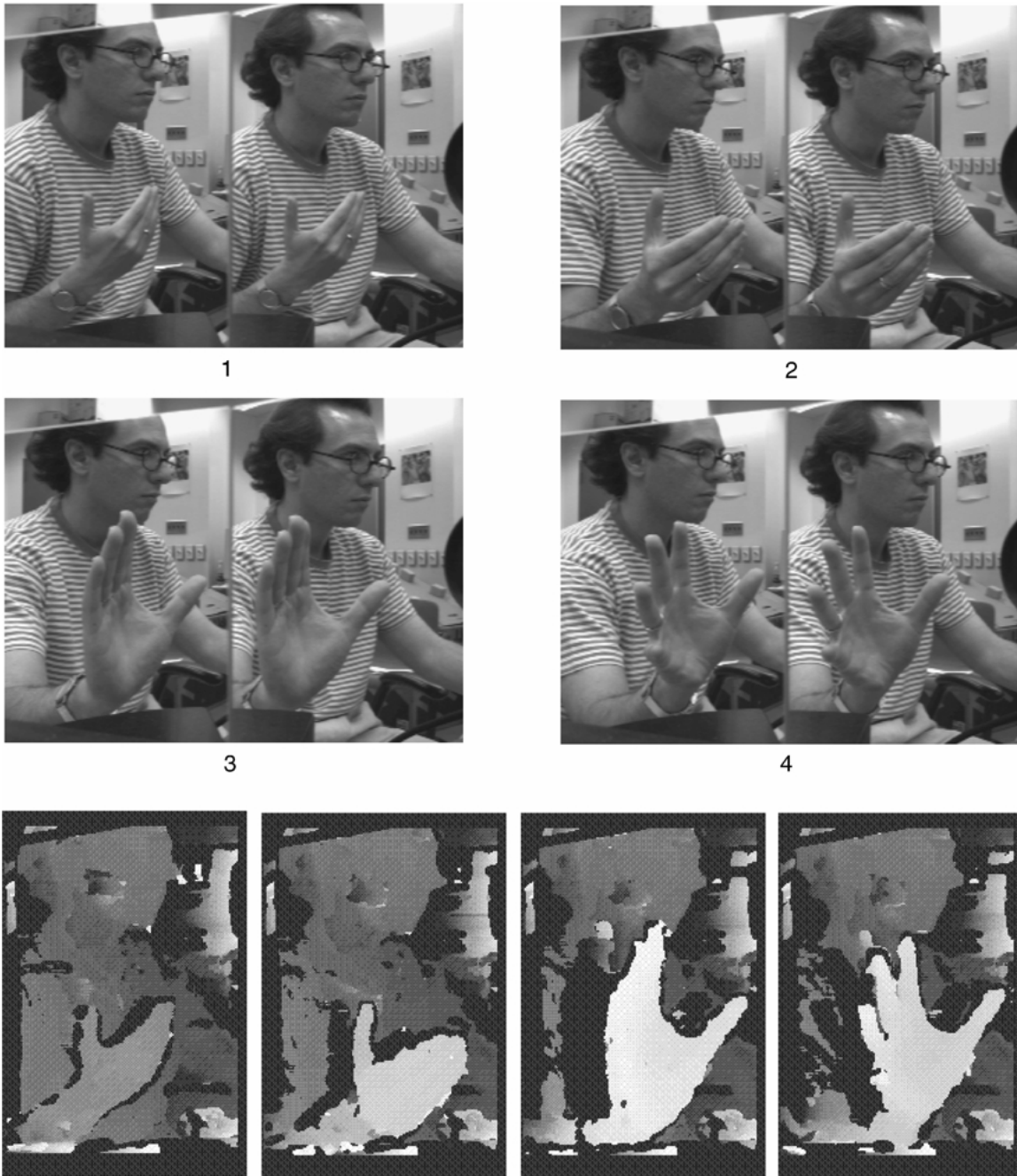


Figure 14. The series of depth maps were generated in real-time from these four catadioptric stereo images.

SAD lends itself to efficient scanline correspondence algorithms.

Stereo matches are found using a standard window based search along scanlines. MMX instructions are used to both compute the SAD and determine the best match. The SAD is computed using MMX parallel

vector instructions. We find the best match through a parallel “tournament” algorithm. The search is limited to an interval of 32 pixels along the epipolar line (scanline) of a  $320 \times 240$  image. In addition we have implemented a left-right checking scheme to prune bad matches. Left-right checking computes a depth

measurement for both a patch in the left image and the patch in the right image it matches. If the two depths are different then no depth is output at that point.

By using the SAD measure, scanline correspondence, and SIMD instructions we are able to achieve a throughput of approximately 20 frames per second on a 300 Mhz Pentium II machine. An example catadioptric stereo image and computed depth map are shown in Fig. 13. Figure 14 shows a sequence of depth maps generated in real-time by our system.

We have examined the geometry of stereo with two planar mirrors in an arbitrary configuration and shown that both the relative orientation and the epipolar geometry are constrained by planar motion. In addition, we have shown how the focal length can be extracted from a single catadioptric image and demonstrated this through a set of experiments. We have also implemented a real-time stereo system which demonstrates that high quality depth maps can be obtained when a single camera is used.

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.

## Acknowledgments

This work was supported in part by DARPA's VSAM Image Understanding Program, under ONR contract N00014-97-1-0553. This paper is an extended version of a paper presented at CVPR '99 (Gluckman and Nayar, 1999).

## Notes

1. Comparison to the Kruppa equations was attempted. However, they gave nonsensical results.
2. [www.inria.fr/robotvis/personnel/zzhang/zzhang-end.html](http://www.inria.fr/robotvis/personnel/zzhang/zzhang-end.html)
3. A sum of squared differences measure (SSD) was also used, however there was no apparent difference.

## References

- Armstrong, M. 1996. Self-calibration from image sequences. Ph.D. Thesis, University of Oxford.
- Armstrong, M., Zisserman, A., and Hartley, R. 1996. Self-calibration from image triplets. In *Proceedings of the 1996 European Conference on Computer Vision*.
- Chaen, A., Yamazawa, K., Yokoya, N., and Takemura, H. 1997. Omnidirectional stereo vision using hyperomni vision. Technical Report 96-122, IEICE (in Japanese).
- Faugeras, O. 1992. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proceedings of the 1992 European Conference on Computer Vision*. Springer-Verlag: Berlin pp. 563–578.
- Faugeras, O., Hotz, B., Mathieu, H., Vieville, T., Zhang, Z., Fau, P., Theron, E., Moll, L., Berry, G., Vuillemin, J., Bertin, P., and Proy, C. 1993. Real-time correlation-based stereo: Algorithm, implementation and application. Technical Report 2013, INRIA Sophia Antipolis.
- Gluckman, J., and Nayar, S.K. 1999. Planar catadioptric stereo: Geometry and calibration. In *Proceedings of the 1999 Conference on Computer Vision and Pattern Recognition*.
- Gluckman, J., Nayar, S.K., and Thoresz, K.J. 1998. Real-time omnidirectional and panoramic stereo. In *Proceedings of the 1998 DARPA Image Understanding Workshop*.
- Goshtasby A., and Gruver, W.A. 1993. Design of a single-lens stereo camera system. *Pattern Recognition*, 26(6):923–937.
- Hartley, R.I. 1992. Estimation of relative camera positions for uncalibrated cameras. In *Proceedings of the 1992 European Conference on Computer Vision*.
- Hartley, R.I. 1995. In defense of the 8-point algorithm. In *Proceedings of the 5th International Conference on Computer Vision*, pp. 1064–1070.
- Hartley, R.I. and Gupta, R. 1993. Computing matched-epipolar projections. In *Proceedings of the 1993 Conference on Computer Vision and Pattern Recognition*.
- Hecht, E. and Zajac, A. 1974. *Optics*. Addison-Wesley: Reading, MA.
- Inaba, M., Hara, T. and Inoue, H. 1993. 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, T., Yoshida, A., Oda, K., Kano, H., and Tanaka, M. 1996. A stereo machine for video-rate dense depth mapping and its new applications. In *Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition*. 1996.
- Kawanishi, T., Yamazawa, K., Iwasa, H., Takemura, H. and Yokoya, N. 1998. Generation of high-resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors. In *International Conference on Pattern Recognition 1998*.
- Konolige, K. 1997. Small vision systems: Hardware and implementation. In *8th Int'l Symposium of Robotics Research*, Hayama, Japan, Oct. 1997
- Longuet-Higgins, H.C. 1981. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135.
- Mathieu, H. and Devernay, F. 1995. Systeme de miroirs pour la stereoscopie. Technical Report 0172, INRIA Sophia-Antipolis (in French).
- Matthies, L. 1993. Stereo vision for planetary rovers: Stochastic modeling to near realtime implementation. *International Journal of Computer Vision*, 8(1):71–91.
- Maybank, S.J. 1993. *Theory of Reconstruction From Image Motion*. Springer-Verlag: Berlin.
- Mitsumoto, H., Tamura, S., Okazaki, K., Kajimi, N., and Fukui, Y. 1992. 3d reconstruction using mirror images based on a plane symmetry recovery method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9):941–945.

- Murray, D.W. 1995. Recovering range using virtual multicamera stereo. *Computer Vision and Image Understanding*, 61(2): 285–291.
- Nayar, S.K. 1988. Robotic vision system. United States Patent 4,893,183.
- Nayar, S.K. and Baker, S. 1997. Catadioptric image formation. In *Proceedings of the 1997 DARPA Image Understanding Workshop*, May 1997.
- Nene, S.A. and Nayar, S.K. 1998. Stereo with mirrors. In *Proceedings of the 6th International Conference on Computer Vision*, Bombay, India, January 1998. IEEE Computer Society.
- Nishimoto Y. and Shirai, Y. 1987. A feature-based stereo model using small disparities. In *cvpr87*, pp. 192–196.
- Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. 1992. *Numerical Recipes in C*. Cambridge University Press.
- Shashua, A. 1998. Omni-rig sensors: What can be done with a non-rigid vision platform? In *Workshop on Applications of Computer Vision*, 1998.
- Southwell, D., Basu, A., Fiala, M., and Reyda, J. 1996. Panoramic stereo. In *Proceedings of the Int'l Conference on Pattern Recognition*, 1996.
- Teoh, W. and Zhang, X.D. 1984. An inexpensive stereoscopic vision system for robots. In *Proc. Int'l Conference on Robotics*, pp. 186–189.
- Triggs, B. 1997. Autocalibration and the absolute quadric. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition*, 1997.
- Tsai, R.Y. 1986. An efficient and accurate camera calibration technique for 3d machine vision.
- Vieville, T. and Lingrand, D. 1995. Using singular displacements for uncalibrated monocular visual systems. Technical Report 2678, INRIA Sophia-Antipolis.
- Xu, G. and Zhang, Z. 1996. *Epipolar Geometry in Stereo, Motion and Object Recognition*. Kluwer: Dordrecht.
- Zeller, C. and Faugeras, O. 1996. Camera self-calibration from video sequences: The kruppa equations revisited. Technical Report 2793, INRIA, Sophia-Antipolis, France.
- Zhang, Z., Deriche, R., Faugeras, O., and Luong, Q.T. 1995. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence Journal*, 78:
- Zhang, Z.Y. and Tsui, H.T. 1998. 3d reconstruction from a single view of an object and its image in a plane mirror. In *International Conference on Pattern Recognition*, 1998.