Rectified Catadioptric Stereo Sensors

Joshua Gluckman, *Member*, *IEEE Computer Society*, and Shree K. Nayar, *Senior Member*, *IEEE*

Abstract—It has been previously shown how mirrors can be used to capture stereo images with a single camera, an approach termed catadioptric stereo. In this paper, we present novel catadioptric sensors that use mirrors to produce rectified stereo images. The scan-line correspondence of these images benefits real-time stereo by avoiding the computational cost and image degradation due to resampling when rectification is performed after image capture. First, we develop a theory which determines the number of mirrors that must be used and the constraints on those mirrors that must be satisfied to obtain rectified stereo images with a single camera. Then, we discuss in detail the use of both one and three mirrors. In addition, we show how the mirrors should be placed in order to minimize sensor size for a given baseline, an important design consideration. In order to understand the feasibility of building these sensors, we analyze rectification errors due to misplacement of the camera with respect to the mirrors.

Index Terms—Stereo imaging, image sensors, stereo rectification.

✦

1 INTRODUCTION

CATADIOPTRIC systems are optical systems that consist of a combination of mirrors and lenses [4]. As demonstrated by several researchers, catadioptrics can be used to design stereo sensors that use only a single camera [11], [3], [5], [9], [15], [12], [2]. Although the systems described by these researchers use a variety of different mirror shapes and configurations, the underlying motivation is the same. By using multiple mirrors, scene points can be imaged from two or more viewpoints while using only a single camera.

Single camera stereo has several advantages over traditional two-camera stereo. Because only a single camera and digitizer are used, system parameters such as spectral response, gain, and offset are identical for the stereo pair. In addition, only a single set of internal calibration parameters needs to be determined. Perhaps most important is that single camera stereo simplifies data acquisition by only requiring a single camera and digitizer and no hardware or software for synchronization.

Real-time stereo systems, whether catadioptric or two-camera, require images to be rectified prior to stereo matching. A pair of stereo images is rectified if the epipolar lines are aligned with the scan-lines of the images. When properly aligned, the search for correspondence is simplified and, thus, real-time performance can be obtained [1]. Once the epipolar geometry of a stereo system is determined, rectifying transformations can be applied to the images [14], [6], [8], [13]. However, rectifying in this manner has two disadvantages for real-time stereo. Applying transformations to the images at runtime is both computationally costly and degrades the stereo data due to the resampling of the images.

An alternative to rectifying the images at runtime is to ensure that the *geometry* of the stereo system produces rectified images. With two-camera stereo this is accomplished by removing any rotation between the two cameras, aligning the direction of translation with the scan lines of the cameras and using identical internal parameters for the two cameras (a difficult task). The geometric requirements of a rectified catadioptric stereo system are not trivial and have not been studied. The purpose of this paper is twofold. First, to develop the constraints that must be satisfied to ensure rectified images and, second, to describe an automated tool for placing the mirrors such that sensor size is minimized for a given baseline. These results can be used to design and build novel compact stereo sensors. In addition, we analyze the effect of misplacement of the camera with respect to the mirrors.

1.1 Previous Work

Several researchers have demonstrated the use of both curved and planar mirrors to acquire stereo data with a single camera. 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 [11], where Nayar suggested a wide field of view stereo system consisting of a conventional camera pointed at two specular spheres (see Fig. 1a). A similar system using two convex mirrors, one placed on top of the other, was proposed by Southewell et al. [15] (see Fig. 1b). Later, Nene and Nayar presented several different catadioptric stereo configurations using a single camera with parabolic, elliptic, and hyperbolic mirrors [12].

Several others have also investigated the use of planar mirrors to design single camera stereo sensors. A sensor designed by Goshtasby and Gruver used two planar mirrors connected by a hinge centered in the field of view of the camera [3]. Gluckman and Nayar demonstrated how two mirrors in an arbitrary configuration can be self-calibrated and used for single camera stereo [2] (see Fig. 1c). Stereo systems using four planar mirrors were proposed by both Inaba et al. [5] and Mathieu and Devernay [9] (see Fig. 1d). By imaging an object and its mirror reflection, a stereo image can also be obtained using only a single mirror [10], [16].

- *The authors are with the Department of Computer Science, Columbia University, 2960 Broadway, New York, NY 10027-6902.*
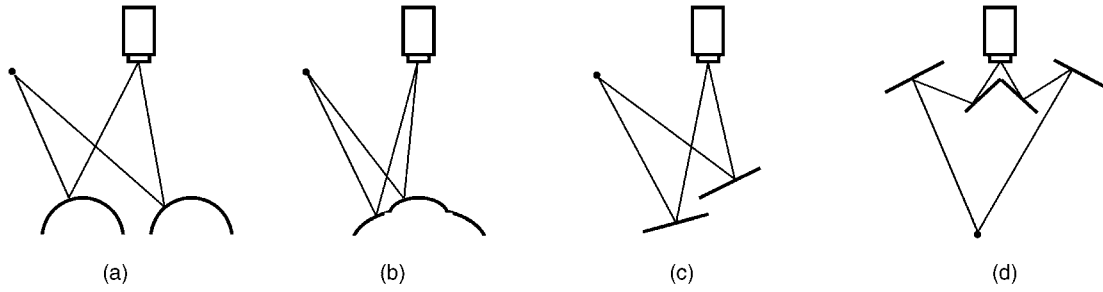 *E-mail: {gluckman, nayar}@cs.columbia.edu.*

Fig. 1. Single camera stereo systems using a variety of mirrors. By imaging two reflections of a scene point, the 3D location can be determined from a single camera. (a) Two spherical mirrors. (b) Two stacked convex mirrors. (c) Two planar mirrors. (d) Four planar mirrors.

In all of these systems, the stereo images are not rectified, therefore, the images must be transformed at runtime prior to stereo matching. One exception is a system described by Lee et al. that uses *prisms* rather than *mirrors* to acquire rectified stereo images from a single camera [7]. Although prisms are an interesting alternative to mirrors it is not clear that compact sensors with sufficient baseline can be designed.

1.2 Background

Before describing the requirements for rectified catadioptric stereo, we review image formation with planar mirrors. As Fig. 2a shows, the image formed when a mirror reflects a scene point is the same perspective image taken by a *virtual camera* located on the opposite side of the mirror. The location of the coordinate system of the virtual camera relative to the coordinate system of the real camera is found by applying a reflection transformation. If we represent the mirror with the normal $\mathbf{n}$ and the distance d measured from the real camera center, the reflection transformation $\mathbf{D}$ is found to be

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

The transformation $\mathbf{D}$ between the real and virtual camera coordinate systems is a combination of a rigid transformation and a switch from a left to a right-handed (or vice-versa) coordinate system. Also, note that a reflection transform is its own inverse:

$$\mathbf{DD} = \mathbf{I}.$$

When two mirrors (see Fig. 2b) reflect a scene point, the virtual camera is found by applying two consecutive reflection transformations. As shown in [2], the resulting transformation represents a planar rigid motion, meaning the direction of translation is orthogonal to the axis of rotation. It was also shown that the axis of rotation is $(\mathbf{n}_1 \times \mathbf{n}_2)$, where $\mathbf{n}_1$ and $\mathbf{n}_2$ are the normals of the two mirrors. In the two mirror case, there is no switch from a left to right-handed system.

For each additional mirror, the virtual camera is found by applying another reflection transformation. In general, if the number of mirrors is odd, then the resulting transformation switches between a left and right handed coordinate system, thus producing a mirror image of the scene.

As shown in Fig. 2c, if the field of view is split such that different mirrors reflect the scene onto different portions of the imaging plane then the scene is imaged from multiple virtual cameras, thus a stereo image is obtained. Now, we will examine how mirrors can be used to obtain a *rectified* stereo image.
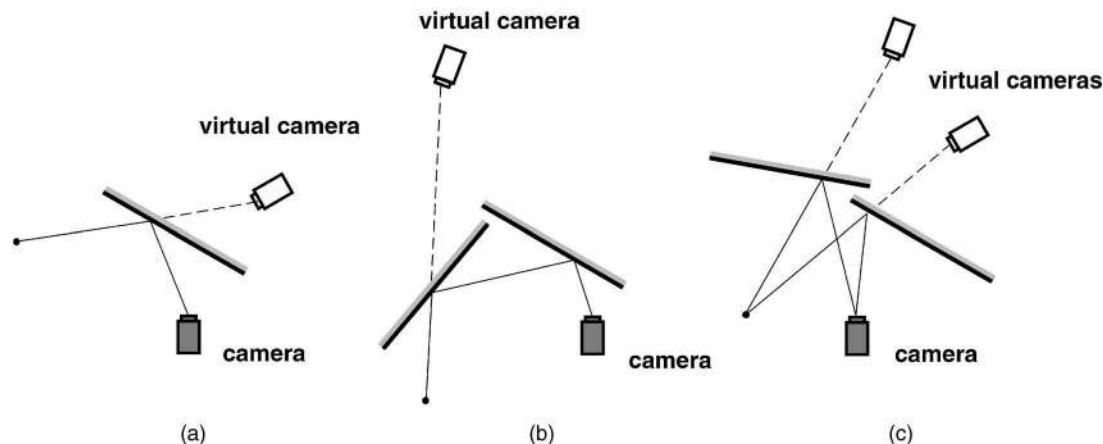

Fig. 2. Image formation with planar mirrors. (a) When a planar mirror reflects a scene point, the image formed is from a virtual camera found by reflecting the real camera about the plane containing the mirror. (b) When two mirrors reflect the scene point, the virtual camera is found by applying two reflections. (c) If the two mirrors produce two reflections of the scene point, then a stereo image is obtained.
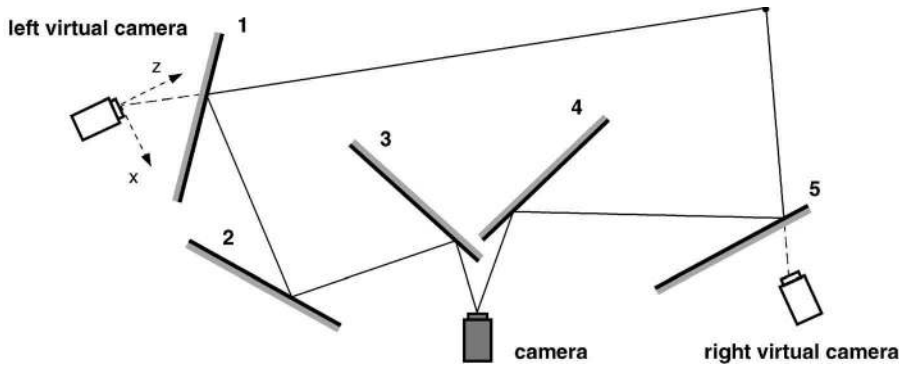
Fig. 3. The relative orientation between the left and right virtual cameras is found by applying consecutive reflection transformations in the order defined by the numbers in the above figure.
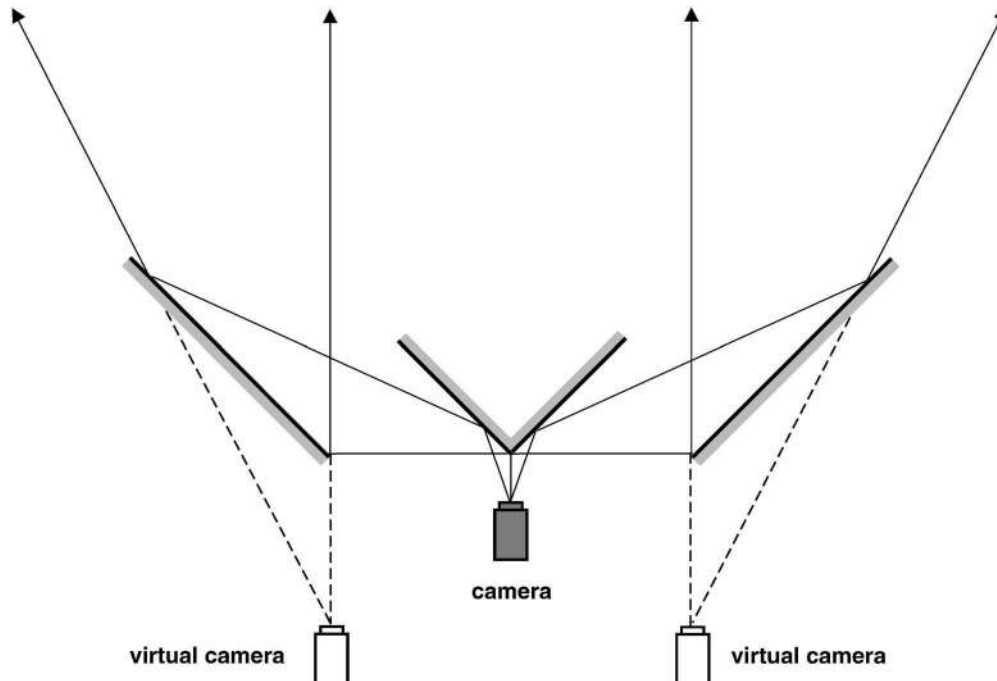

Fig. 4. When two sets of parallel mirrors are used a rectified stereo system can be constructed. However, this solution is not practical because the two virtual cameras do not share a common field of view.

2   HOW MANY MIRRORS ARE NEEDED?

To produce rectified images, a stereo system must meet several requirements. There must be no relative rotation between the two cameras, the translation must be parallel to the scan-lines of the image plane, and the internal parameters of the two cameras must be identical.[1] For catadioptric stereo, the last requirement is met because only a single camera is used. To ensure the first two requirements, the mirrors must satisfy

$$\mathbf{D}_1 \dots \mathbf{D}_i \dots \mathbf{D}_m = \begin{pmatrix} 1 & 0 & 0 & b \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad (1)$$

where b is the baseline, m is the number of mirrors used, and $\mathbf{D}_i$ is the reflection transformation produced by the ith mirror. The mirrors are ordered as shown in Fig. 3 and each

1. This is assuming radial distortions are negligible.

mirror is defined in a coordinate system attached to the left virtual camera, where the x-axis is along the scan-lines and the z-axis is in the direction of the optical axis of the left virtual camera.

Although satisfying (1) is sufficient to ensure rectification, there is one problem. Because we split the field of view of the real camera between the two virtual cameras, we must guarantee that the fields of view properly overlap in order to have a meaningful stereo system. Fig. 4 shows a four mirror system where the two virtual cameras are rectified but do not share a common field of view. When the field of view is split between two different systems of mirrors, each virtual camera receives only half of the field of view of the real camera. Clearly, in practice the two half fields of view must overlap.

As shown in Fig. 5, each virtual camera receives either the left or right half field of view. When rectified, there are four possible configurations, depending upon the number of reflections. However, only the bottom two configurations lead to a practical stereo system for most applications. The
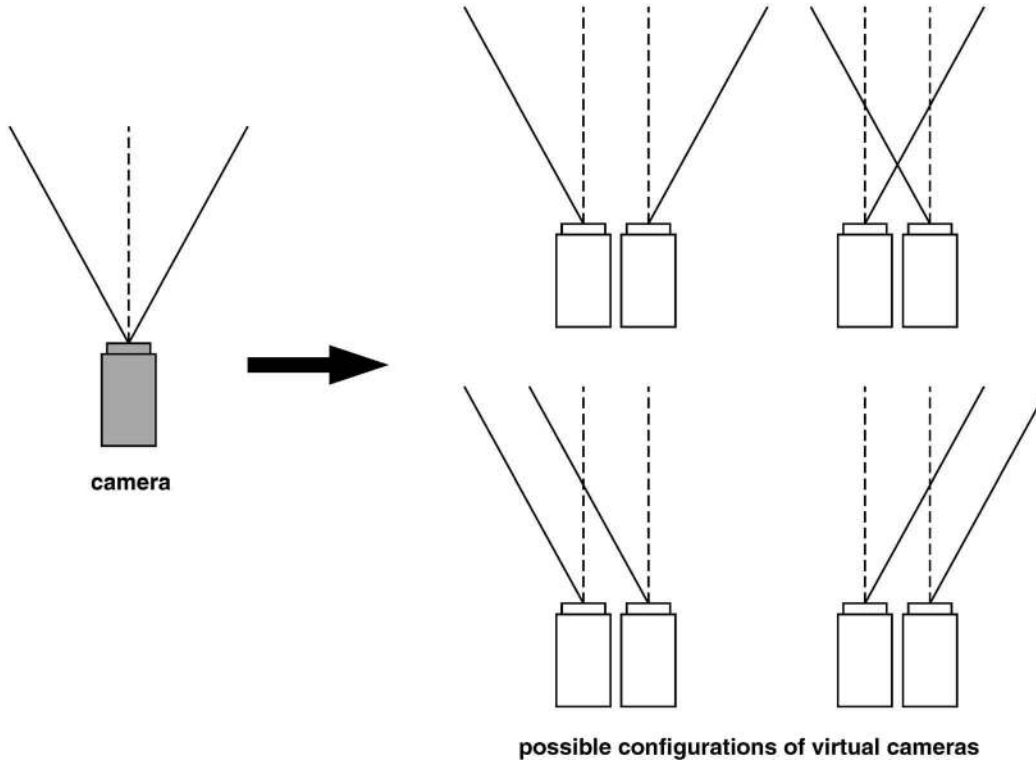
Fig. 5. Once rectified, there are four possible configurations of the virtual cameras. However, only the bottom two lead to a practical solution. The top left configuration has no overlapping field of view and the common field of view of the top right configuration is only a narrow beam. To obtain one of the bottom configurations an odd number of reflections must be used.

top left configuration has no overlapping field of view and the top right only sees a narrow beam the width of the baseline. Although, for applications where the scene of interest lies close to the sensor, this configuration may be useful. To obtain one of the configurations on the bottom, one half field of view must be reflected relative to the other and therefore an odd number of reflections are required. We enforce this by changing the direction of the x-axis

$$\mathbf{D}_1 \ldots \mathbf{D}_i \ldots \mathbf{D}_m = \begin{pmatrix} -1 & 0 & 0 & b \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \qquad (2)$$

It is straightforward to show that for any number of odd reflections a solution exists. For $m = 1$, the mirror normal $\mathbf{n}_1 = [1, 0, 0]^T$ leads to a solution. For all odd m, $(m > 1)$ a trivial solution can be obtained by adding $\frac{m-1}{2}$ pairs of identical reflection transformations; because a reflection transformation is its own inverse, each pair will cancel out. Although there are many solutions to (2), most are not physically realizable due to occlusions and intersecting mirrors. Next, we will discuss possible solutions using one and three mirrors. Five or more mirrors can be used; however, these systems are complex and their advantages are unclear.

3 SINGLE MIRROR RECTIFIED STEREO

To obtain a rectified image with a single mirror, the plane containing the mirror must satisfy

$$\mathbf{D}_1 = \begin{pmatrix} \mathbf{I} - 2\mathbf{n}_1\mathbf{n}_1^T & 2d_1\mathbf{n}_1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 & b \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad (3)$$

where $\mathbf{n}_1$ and d_1 are the normal and distance of the mirror from the camera coordinate system. For this to be satisfied, $\mathbf{n}_1 = [1, 0, 0]^T$. Thus, the only solution using a single mirror occurs when the normal of the mirror is parallel to the scan-lines of the camera (the x-axis) as shown in Fig. 6. The stereo system will remain rectified for any distance d_1; however, the baseline b will change as $b = 2d_1$.

The advantage of this solution is its simplicity. However, because a finite mirror must be used, the field of view of the virtual camera is limited by the angle the mirror subtends with respect to the virtual camera (see Fig. 6). The field of view α is related to the baseline b and the length of the mirror h as

$$\alpha = \arctan\left(\frac{2h}{b}\right) - \frac{\pi}{2} + \frac{\beta}{2}. \qquad (4)$$

Therefore, if a large baseline is desired then a large mirror must be employed or else the field of view will be severely limited.

Because the mirror is finite, a field of view asymmetry exists between the real and virtual cameras. A larger portion of the image detector is used by the real camera. As shown in Fig. 7, this asymmetry can be removed by shifting the image detector with respect to the center of projection of the camera. For applications where the scene of interest lies close to the camera, this has the benefit of increasing the viewing volume close to the stereo system.
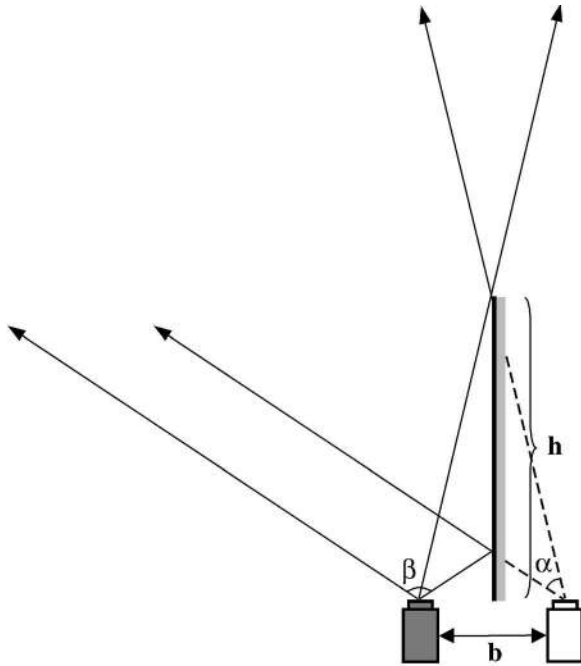
stereo four constraints between the mirrors and the camera must be satisfied. From (2), we know that the mirrors must be placed such that

$$\mathbf{D}_1\mathbf{D}_2\mathbf{D}_3 = \begin{pmatrix} -1 & 0 & 0 & b \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{5}$$

Using (5), we will first derive three constraints by determining how the mirrors must be placed so that there is no rotation between the virtual cameras. Then, we will show the remaining constraint that arises from requiring the direction of translation to be parallel to the scan-lines (the x-axis).

4.1 Rotation Constraints

The upper left 3×3 block of the matrix in (5) refers to the mutual rotation between the two virtual cameras. For rectification, we have:

$$\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{6}$$

where

$$\mathbf{R}_i = \left[\mathbf{I} - 2\mathbf{n}_i\mathbf{n}_i^T \right]. \tag{7}$$

Because each $\mathbf{R}_i$ is both symmetric and its own inverse, (6) can be written as

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{R}_3 = \mathbf{R}_1\mathbf{R}_2. \tag{8}$$

Fig. 6. To obtain a rectified image with a single mirror, the normal of the mirror must be parallel to the scan-lines of the camera. Note that the field of view of the right virtual camera will be limited by the finite size of the mirror.

4 THREE MIRROR RECTIFIED STEREO

We can overcome the limitations of single mirror rectified stereo by incorporating additional mirrors. With three mirrors, we can ensure that the field of view is equally shared between the two virtual cameras (see Fig. 8). Furthermore, a large baseline can be obtained using relatively small mirrors. However, we cannot arbitrarily place the three mirrors. As we will show, to obtain rectified

image detector
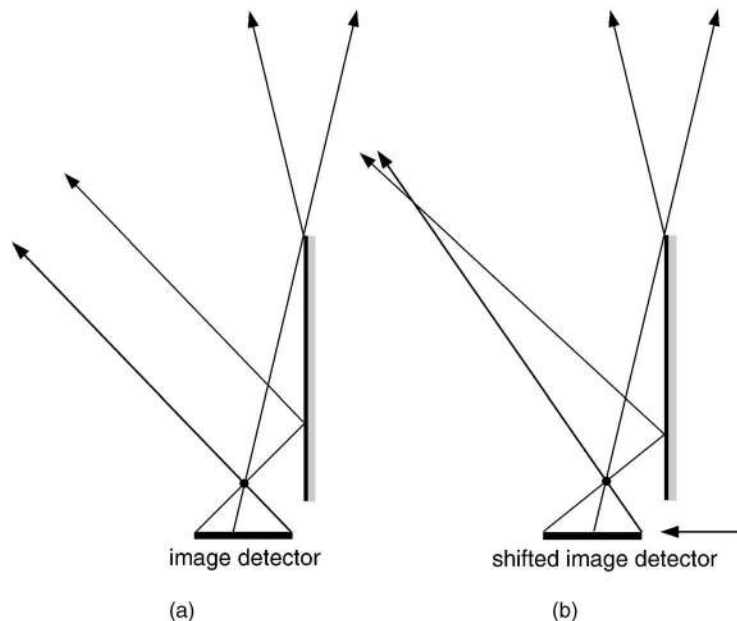(a)

shifted image detector
(b)

Fig. 7. Shifting the image detector to reduce asymmetry in the stereo field of view. (a) When a single mirror is used the left camera view uses a larger portion of the image detector than the right camera view. (b) The asymmetry can be removed by shifting the image detector with respect to the imaging lens (center of projection).
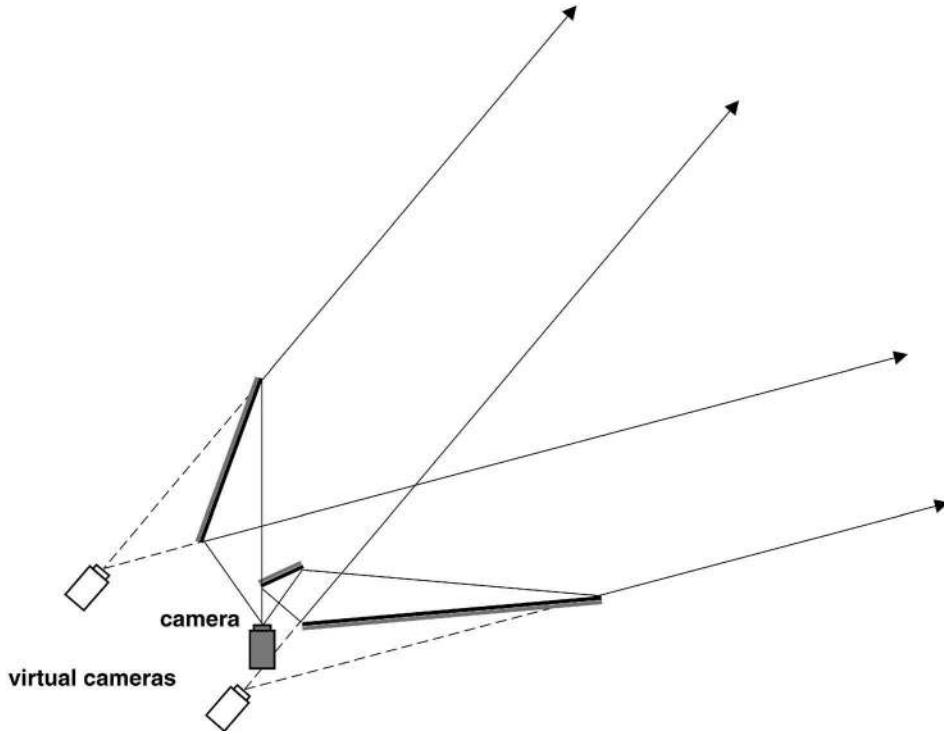
Fig. 8. Three mirror rectified stereo. By using three mirrors a rectified stereo system can be designed where the field of view of each virtual camera is half the field of view of the real camera.

Recall that $\mathbf{R}_1\mathbf{R}_2$ is a rotation matrix with a rotational axis orthogonal to both the mirror normals, which is $\mathbf{n}_1 \times \mathbf{n}_2$. This implies that

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{R}_3(\mathbf{n}_1 \times \mathbf{n}_2) = (\mathbf{n}_1 \times \mathbf{n}_2). \qquad (9)$$

It is straightforward to show that

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{R}_3\left(\mathbf{n}_3 \times [1,0,0]^T\right) = \left(\mathbf{n}_3 \times [1,0,0]^T\right) \quad (10)$$

must also hold. Therefore, to satisfy (6), it is necessary that both (9) and (10) are satisfied, meaning

$$(\mathbf{n}_1 \times \mathbf{n}_2) = \lambda\left(\mathbf{n}_3 \times [1,0,0]^T\right). \qquad (11)$$

The scale factor λ can be removed by taking the dot product with $\mathbf{n}_1$ or $\mathbf{n}_3$, giving us two independent constraints,

$$\mathbf{n}_1 \cdot \left(\mathbf{n}_3 \times [1,0,0]^T\right) = 0 \qquad (12)$$

and

$$\mathbf{n}_3 \cdot (\mathbf{n}_1 \times \mathbf{n}_2) = 0. \qquad (13)$$

This implies that the normals of the three mirrors $\mathbf{n}_1$, $\mathbf{n}_2$, and $\mathbf{n}_3$ and the x-axis are all co-planar. When the normals are coplanar, the mirrors rotate the virtual cameras about a common axis. Of course, we still need to orient the mirrors so that the rotational angle is cancelled. But, now we can simplify our analysis by working in the two dimensions of the plane containing the normals and the x-axis. In 2D, the

mirrors are represented by lines where θ_i is the angle the ith mirror normally makes with the x-axis of the real camera. For simplicity, we have moved the x-axis to the real camera.

If we consider a ray leaving the camera center at angle α and a corresponding ray leaving the camera at an angle $\pi - \alpha$, then the mirrors must be angled such the two reflected rays are parallel, thus ensuring there is no rotation between the virtual cameras (see Fig. 9). After being reflected by the mirror oriented at angle θ_1, the angle of the left ray is

$$\alpha_1 = 2\theta_1 - \alpha. \qquad (14)$$

The angle of the right ray α_2 after reflection by the two mirrors is

$$\alpha_2 = 2\theta_3 - 2\theta_2 - \alpha. \qquad (15)$$

The two rays are parallel if $\alpha_1 = \alpha_2$. Therefore, the rotation is cancelled if

$$\theta_3 - \theta_2 = \theta_1. \qquad (16)$$

We can express (16) in terms of the normals of the mirrors as,

$$\mathbf{n}_3 \cdot \mathbf{n}_2 = \mathbf{n}_1 \cdot [1,0,0]^T. \qquad (17)$$

To summarize, if the normals of the mirrors satisfy the three constraints (12), (13), and (17), then there will be no rotation between the two virtual cameras.

4.2 Translation Constraints

So far, we have three constraints that must be satisfied to rectify the virtual cameras. However, we also need to
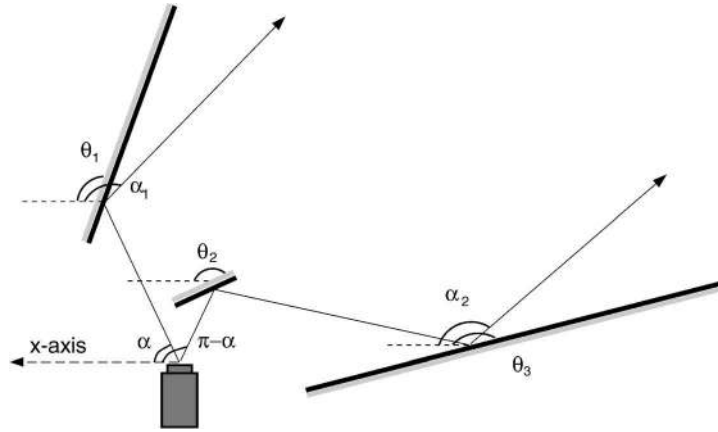
Fig. 9. The mirrors must be angled such that two rays at angles α and $\pi - \alpha$ are parallel after being reflected by the mirrors, that is $\alpha_1 = \alpha_2$. Doing so will ensure that there is no rotation between the two virtual cameras.

ensure that the direction of translation between the virtual cameras is along the x-axis. For this, we will need to examine the translational part of (5), that is

$$2d_3\mathbf{R}_1\mathbf{R}_2\mathbf{n}_3 + 2d_2\mathbf{R}_1\mathbf{n}_2 + 2d_1\mathbf{n}_1 = \begin{pmatrix} b \\ 0 \\ 0 \end{pmatrix}. \qquad (18)$$

After multiplying throughout with $\mathbf{R}_3\mathbf{R}_2\mathbf{R}_1$ and substituting (6), we get

$$2d_3\mathbf{R}_3\mathbf{n}_3 + 2d_2\mathbf{R}_3\mathbf{R}_2\mathbf{n}_2 + 2d_1 \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{n}_1 = \begin{pmatrix} -b \\ 0 \\ 0 \end{pmatrix}. \qquad (19)$$

Next, by substituting (7) for the $\mathbf{R}_i$ we have

$$-2d_3\mathbf{n}_3 - 2d_2\mathbf{n}_2 + 4(\mathbf{n}_3 \cdot \mathbf{n}_2)d_2\mathbf{n}_3 + 2d_1 \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{n}_1 = \begin{pmatrix} -b \\ 0 \\ 0 \end{pmatrix}. \qquad (20)$$

Although satisfying this equation ensures rectification, there is only one independent constraint in (20). Since the baseline b is arbitrary, we expand only the last two equations which are

$$-n_{3y}d_3 + \left(2(\mathbf{n}_3 \cdot \mathbf{n}_2)n_{3y} - n_{2y}\right)d_2 + n_{1y}d_1 = 0 \qquad (21)$$

and

$$-n_{3z}d_3 + (2(\mathbf{n}_3 \cdot \mathbf{n}_2)n_{3z} - n_{2z})d_2 + n_{1z}d_1 = 0, \qquad (22)$$

where $\mathbf{n}_i = [n_{ix}, n_{iy}, n_{iz}]^T$. Note that the vector equation (20) is composed of the normals of the mirrors and the x-axis. Since the rotation constraints ensure these vectors are coplanar, we define the coordinate system such that these vectors lie in the x-z plane. When the normals are in the x-z plane, then $n_{iy} = 0$ for all i and (21) is satisfied, leaving only one constraint (22).

We have now derived the four constraints for three mirror rectified stereo. To summarize, if the normals of the three mirrors are coplanar with the x-axis, the angles between the mirrors satisfy (17) and the distances to the mirrors are chosen such that (22) is satisfied, then the two virtual cameras will be rectified.

5 COMPACT SENSOR DESIGN

For many stereo applications, the compactness of the sensor is important. In this section, we describe an automated tool for catadioptric stereo sensor design. Given design parameters such as baseline, field of view, and size of the real camera, we compute the optimal placement of the mirrors such that the virtual cameras are rectified and sensor size is minimized.

To simplify the optimization, we assume the plane containing the normals of the mirrors is the x-z plane, meaning there is no tilt between the camera and the mirrors. Now, each of the three mirrors is represented by a line, so there are only six parameters: θ_1, θ_2, θ_3, d_1, d_2, and d_3. Because we are restricted to the x-z plane, (12) and (13) are satisfied, thus there are only two rectification constraints, one on the angles of the mirrors (16) and one on the distances (22). Given a desired baseline b, we have one more constraint from (20)

$$2n_{3x}d_3 + (2n_{2x} - 4(\mathbf{n}_3 \cdot \mathbf{n}_2)n_{3x})d_2 + 2n_{1x}d_1 = b. \qquad (23)$$

Three constraints on six parameters leaves three free parameters. To optimize these parameters, some criteria for sensor size must be chosen. One simple measure is the perimeter of the bounding box of the mirrors and the center of projection of the camera (see Fig. 10).

To find the best configuration, we search through all possible locations (θ_1, d_1) of the first mirror and possible angles θ_2 for the second mirror. The remaining three parameters, the distance d_2 to the second mirror, the distance d_3 to the third mirror, and the angle θ_3 of the third mirror are found by solving (16), (22), and (23), where $n_{ix} = \cos\theta_i$ and $n_{iz} = \sin\theta_i$. For each set of computed parameters, we determine the end points of the mirrors by tracing the optical axis and the limiting rays of the field of view and intersecting them with the mirrors. Once the
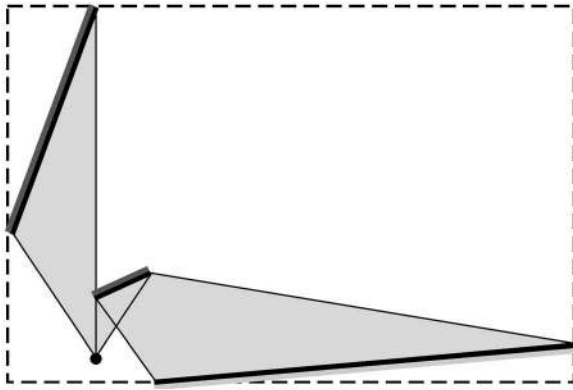
Fig. 10. We search for the mirror locations that satisfy the rectification constraints and minimize the perimeter of the bounding box, for a given baseline and field of view. The bounding box is found by intersecting the field of view of the camera with the mirror locations and determining the extreme points among the mirror end points and camera center of projection.

end points are found, the perimeter of the bounding box is computed. The design that minimizes the perimeter is chosen.

We only admit solutions where the mirrors do not occlude each other. Thus, we ensure that the ray r_2, in Fig. 11, does not intersect the mirrors M_2 and M_3 and the ray r_3 does not intersect M_2. In addition, we only consider solutions where the real camera does not see itself and, thus, we ensure the ray r_1 is a minimum distance c from the camera center of projection (note that c is proportional to the baseline b and determined from the size of the camera).

Fig. 12 shows two optimal configurations for a camera with a $60°$ field of view. As shown in Fig. 12a, when $c = 0$, the second mirror is infinitesimally small and located at the center of projection of the camera. Fig. 12b shows the optimal solution when $c = 0.1$. As both c and the field of

view increase, the optimal sensor size also increases. Fig. 13 shows a sampling of optimized sensors for a variety of different fields of view and values for c. Note that we may use the same approach to find the optimal sensor for some other size criterion, such as the area of the bounding box.

6 ERROR TOLERANCE

Once a mirror configuration is chosen, the location and orientation of the camera is determined up to a translation orthogonal to the mirror normals. Although the mirrors can be placed with high accuracy, inevitably errors will be introduced when positioning the camera and lens with respect to the mirrors because it is difficult to determine the precise location of the center of projection and optical axis of an imaging system. In this section, we examine the effect of these placement errors on the rectified geometry.

For simplicity, we assume the desired configuration is one where the optical axis of the camera is in the plane defined by the mirror normals and the center of projection of the camera, referred to as the mirror normal plane. We will look at three sources of errors: misplacement of the center of projection of the imaging system, misalignment of the orientation of the optical axis within the mirror normal plane, and tilt of the optical axis with respect to the mirror normal plane.

6.1 Misplacement of the Center of Projection

Misplacement of the camera center of projection does not alter the normals of the mirrors; therefore, we only need to consider the effect on the location and not the orientation of the virtual cameras. Moving the center of projection of the camera from the desired location c to the point $\hat{c}$ causes the virtual cameras v and v' to shift to new locations $\hat{v}$ and $\hat{v}'$ (see Fig. 14). If D_1, D_2, and D_3 are reflection transforma-
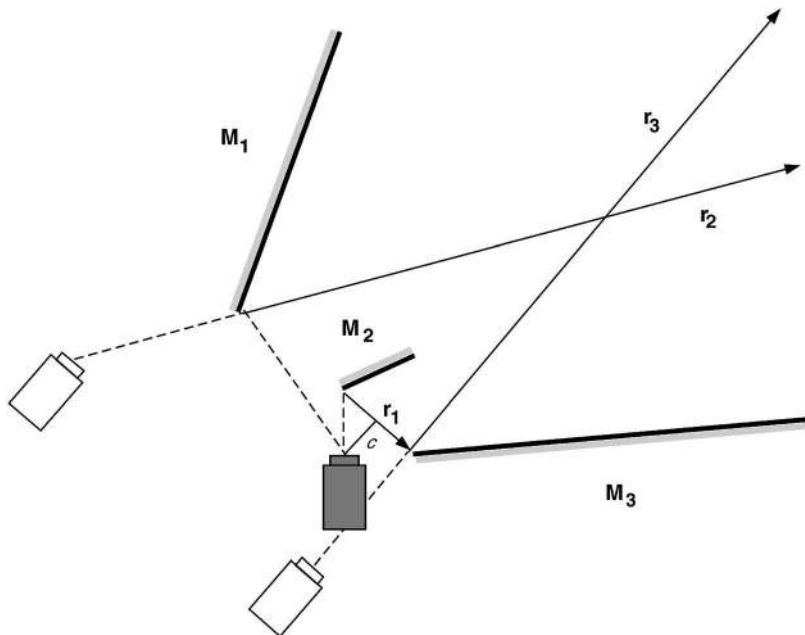

Fig. 11. It is important that the reflected scene rays are not occluded by the mirrors. Therefore, we ensure that the ray r_2 does not intersect the mirrors M_2 and M_3 and the ray r_3 does not intersect M_2. If we have some notion of the size of our camera, then we can also ensure that the ray r_1 is at a minimum distance c from the camera center of projection.
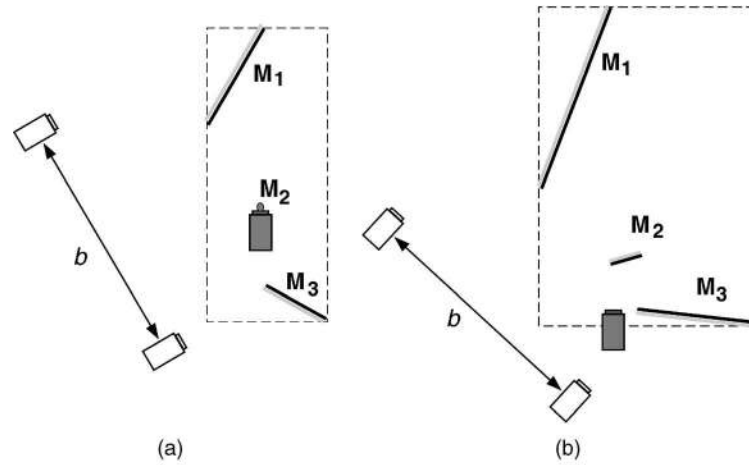
Fig. 12. Optimized stereo sensors for baseline b. (a) If we do not consider the size of the camera, $c = 0$, then the optimal solution has one mirror at the center of projection. (b) This configuration is optimal when $c = 0.1$. For scale, the baseline between the virtual cameras is shown.
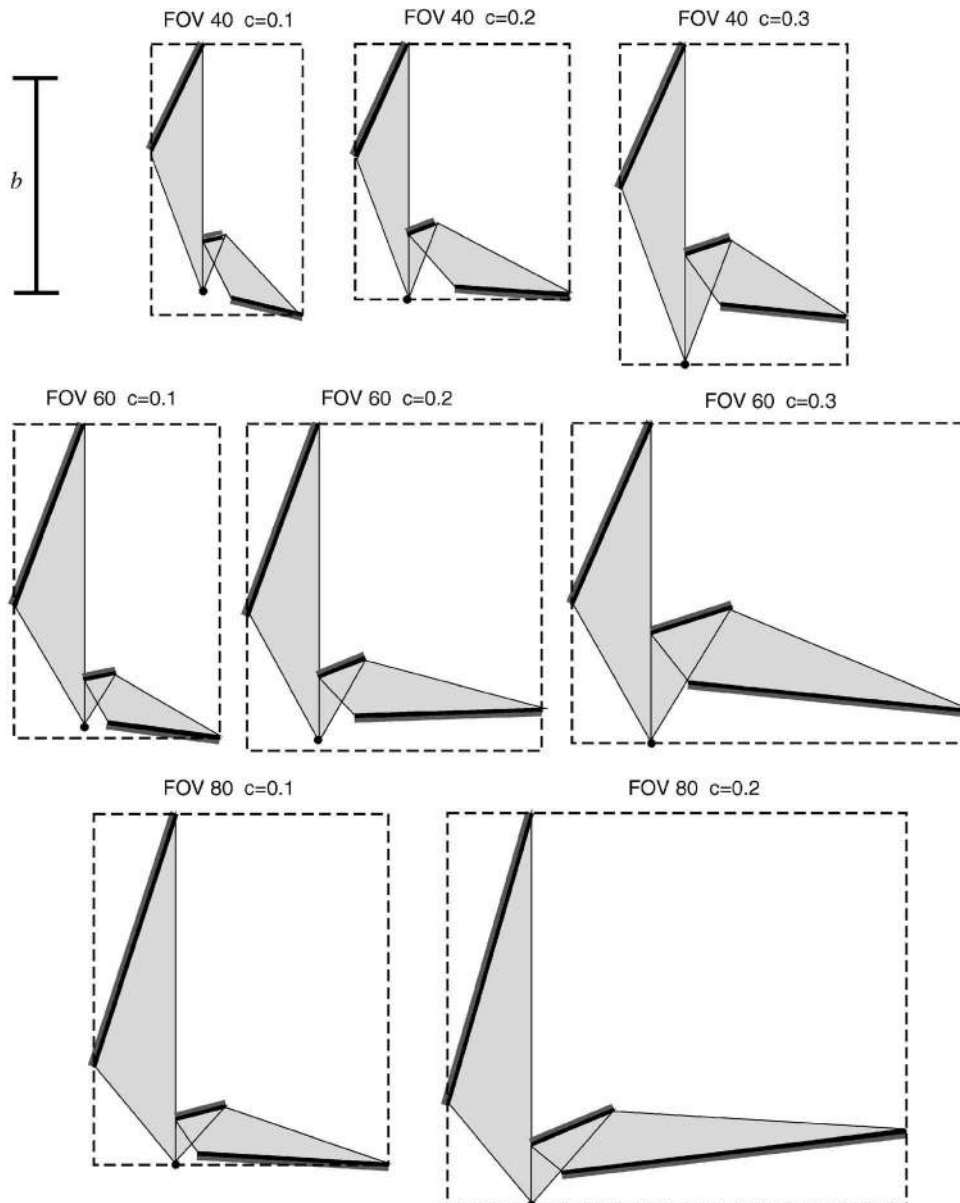

Fig. 13. The size of the optimized sensor depends on the field of view (FOV) and the value for c. In the upper left corner, the baseline b, which is the same for all the sensors, is shown for scale.
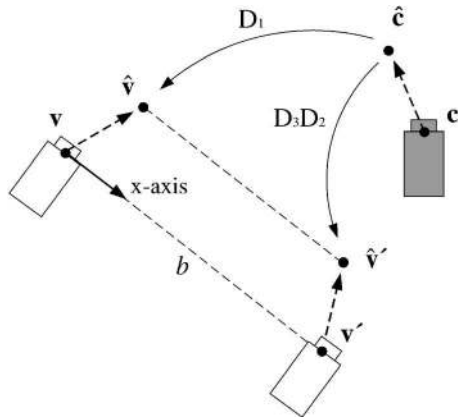
Fig. 14. Moving the center of projection of the camera from c to $\hat{c}$ causes the virtual cameras v and v' to shift to the points $\hat{v}$ and $\hat{v}'$. The location of $\hat{v}$ is found by reflecting $\hat{c}$ about the first mirror using the transformation $\mathbf{D}_1$, and $\hat{v}'$ is found by reflecting $\hat{c}$ about the second and third mirrors using $\mathbf{D}_3\mathbf{D}_2$. Although the baseline is no longer b, the new virtual cameras remain rectified because the direction of translation is still along the x-axis.

tions about the three mirrors, then $\hat{v} = \mathbf{D}_1\hat{c}$ and $\hat{v}' = \mathbf{D}_3\mathbf{D}_2\hat{c}$. Since a reflection transformation is its own inverse, we also have $\hat{v}' = \mathbf{D}_3\mathbf{D}_2(\mathbf{D}_1\mathbf{D}_1)\hat{c}$ which leads to $\hat{v}' = \mathbf{D}_3\mathbf{D}_2\mathbf{D}_1\hat{v}$. From the rectification constraint (5), it is easy to show that $\hat{v}'$ is related to $\hat{v}$ by just a shift along the x-axis. Thus, the rectification of the virtual cameras is *invariant* to the position of the center of projection of the real camera, the only effect being a change in the baseline. This is an important property because it is difficult to accurately locate the center of projection of an imaging system.
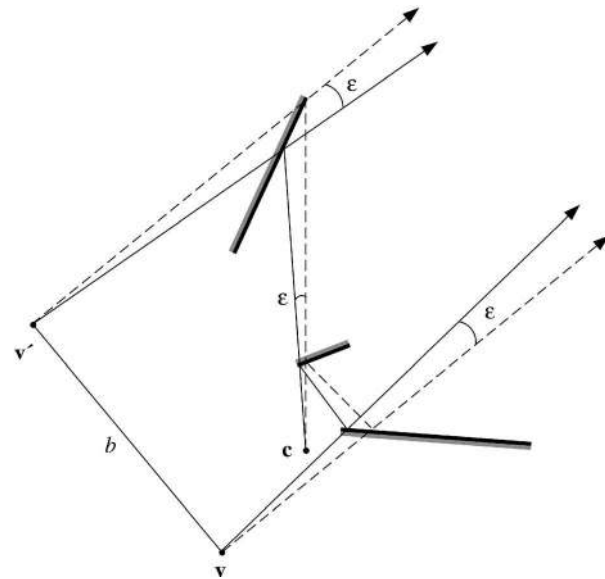

Fig. 15. The dotted lines shows the desired orientation of c the camera and the rectified virtual cameras v and v'. The solid lines show the orientation of the virtual cameras if the real camera has an angular error of ε. In this case, the virtual cameras verge and the cameras will no longer be rectified. The amount of vergence and, thus, the rectification error is independent of the configuration of the mirrors.

6.2 Misalignment of the Optical Axis

Next, we consider misalignment of the orientation of the camera within the mirror normal plane. Misaligning the optical axis of the camera does not alter the locations of the virtual cameras but only their orientation. As shown in Fig. 15, an angular error of ε causes the virtual cameras to verge resulting in a rotation between the virtual cameras.
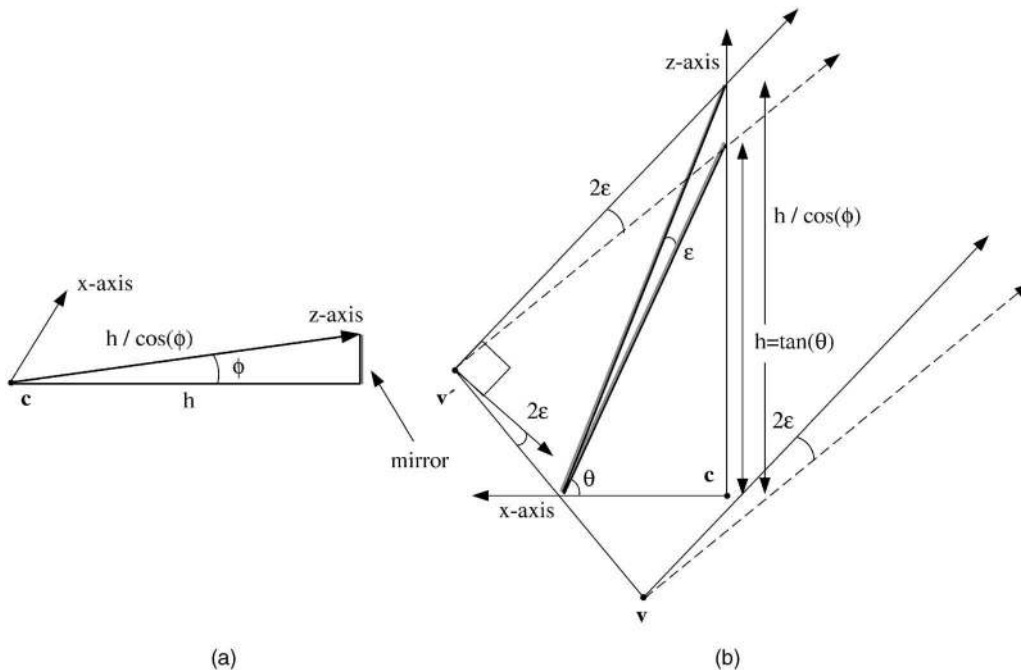

Fig. 16. (a) shows a side view of the camera located at c with the optical axis tilted and pointing along the z-axis. Tilting the z-axis by angle ϕ causes the distance to the mirror along the optical axis to increase from h to $h/\cos\phi$. (b) shows a top view. When the camera is tilted, the angle of the mirror changes with respect to the coordinate system of the camera. Although the virtual cameras remain parallel, they both rotate by an angle 2ε where ε is the amount of the change in the angle of the mirror.
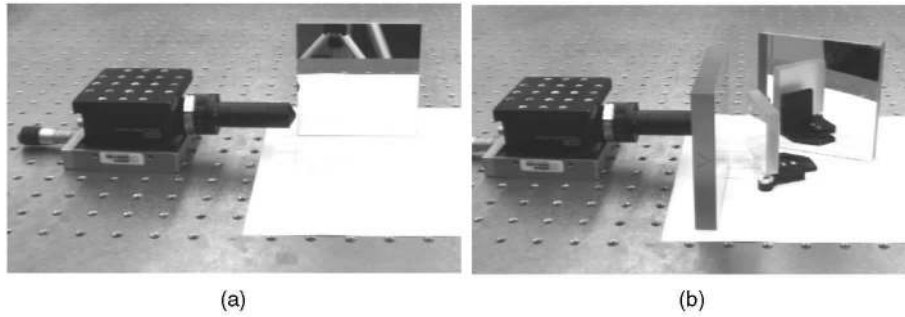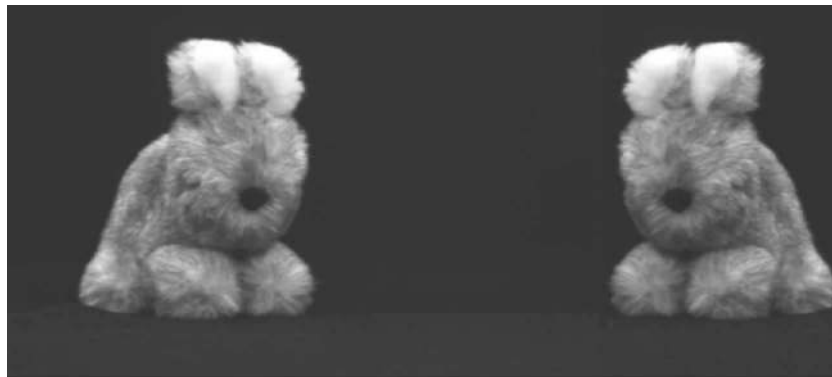
Fig. 17. One and three mirror rectified stereo systems using a Sony XC-75 camera with a Computar 4mm pinhole lens. (a) A single mirror is placed such that the mirror normal is perpendicular to the optical axis of the imaging system. (b) Three mirrors are placed in a rectified configuration.

Note that the amount of vergence is independent of the particular configuration of the mirrors. For a verged system, the rectification error in pixel units depends on the amount of vergence, the field of view of the camera, and the location in the image. The rectification error is subpixel everywhere in the image if $\varepsilon < \cot^{-1}\left(\frac{P}{2}\tan\frac{\alpha}{2}\right)$, where P is the height of the image in pixels and α is the field of view of the camera. When the field of view is less than $90°$, a vergence angle of about $0.2°$ can be tolerated at a typical imaging resolution of approximately 500 pixels.

6.3 Tilt of the Optical Axis

Another source of error is the tilt of the camera with respect to the mirror normal plane. Tilting the camera causes the mirror normals to rotate about the x-axis. Given the rotation $\mathbf{R}_x$, the new mirror normals are $\hat{\mathbf{n}}_1 = \mathbf{R}_x\mathbf{n}_1$, $\hat{\mathbf{n}}_2 = \mathbf{R}_x\mathbf{n}_2$, and $\hat{\mathbf{n}}_3 = \mathbf{R}_x\mathbf{n}_3$. It is straightforward to show that, if the angle constraint $\mathbf{n}_3 \cdot \mathbf{n}_2 = \mathbf{n}_1 \cdot [1,0,0]^T$, given in (17), holds for the old normals, then $\hat{\mathbf{n}}_3 \cdot \hat{\mathbf{n}}_2 = \hat{\mathbf{n}}_1 \cdot [1,0,0]^T$ also holds for the new normals. Thus, the virtual cameras remain parallel when the camera is tilted.
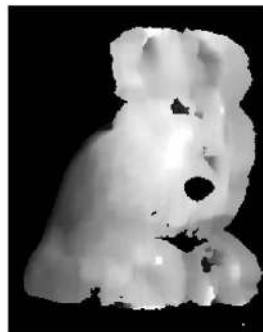
Fig. 18. Single mirror rectified stereo. (a) An image (cropped for display) captured by the single mirror stereo system. Note that the right side of the image is reflected. To demonstrate that the images are rectified, we perform stereo matching along the scan-lines of the image after removing the reflection. (b) and (c) are the depth maps computed using SSD and normalized cross-correlation with a 15×15 window. Normalized cross-correlation performs slightly better due to the intensity differences introduced by the reflection of scene rays at acute angles with the mirror. Depth is not computed for background pixels.
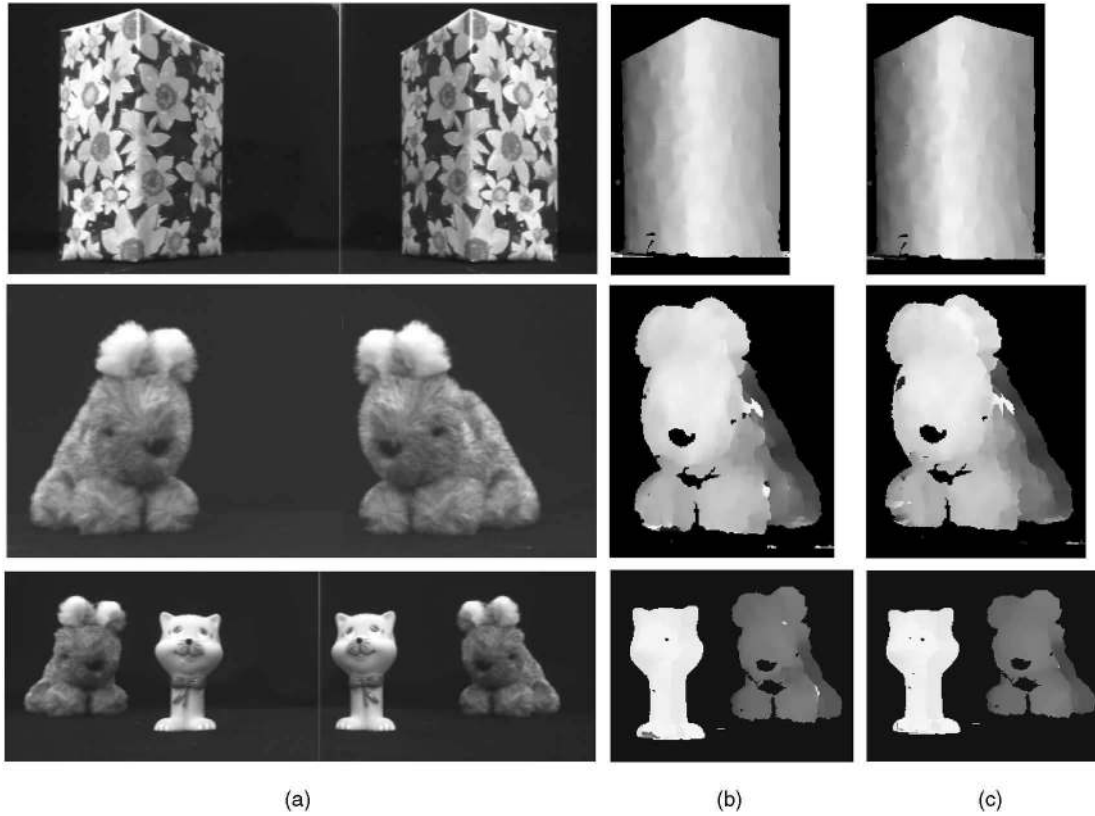
Fig. 19. Images and depth maps using three mirror rectified stereo. (a) Three 640×480 images (cropped for display) captured using three mirrors. Before stereo matching along the scan-lines is performed the right half of each image is flipped to remove the reflection. (b) and (c) are the depth maps obtained using SSD and normalized cross-correlation with a 15×15 window. Note that the difference between the two is negligible thus emphasizing the fact that when only a single camera is used SSD is sufficient for stereo matching. Depth is not computed for the background pixels.

Although the virtual cameras remain parallel, they both rotate causing the direction of translation to turn away from the scan-lines. The change in the direction of translation is dependent on ϕ the amount of tilt, but, unlike the previous case, this error is dependent on the angles of the mirrors. Fig. 16 shows the effect on a single mirror of tilting the camera. Tilting the camera causes the intersection of the mirror with the optical axis to change and thus the angle of the mirror with respect to the coordinate system of the camera. If the camera is tilted by ϕ and the angle of the mirror is given by θ, then ε the change in the angle of the mirror is

$$\varepsilon = \tan^{-1}\left(\frac{\tan\theta}{\cos\phi}\right) - \theta. \qquad (24)$$

For a given tilt, ε is minimized as the mirror angle approaches either $0°$ or $90°$ and is maximum when $\theta = 45°$. The rectification error depends on the error in the direction of translation, which is 2ε. Even when $\theta = 45°$, this error grows slowly as $\frac{\partial\varepsilon}{\partial\phi} = 0$ at $\phi = 0$. For example, a tilt error of $\phi = 1°$ results in a direction of translation error of less than $0.009°$. Therefore, the rectification error is much less sensitive to tilt errors than to misalignment errors. In this analysis, we have assumed the desired configuration is one where the optical axis is in the mirror normal plane. If the desired configuration is one where the camera points up or down at oblique angles with respect to the mirrors, the

rectification will be more sensitive to small tilt errors. In summary, misplacement of the camera center does not effect the rectified geometry, misaligning the optical axis introduces vergence between the virtual cameras, and tilting the optical axis rotates the cameras causing a change in the direction of translation.

7 EXPERIMENTAL RESULTS

We have used the constraints introduced in the preceding sections to build both one and three mirror stereo sensors. As in Fig. 17a, we can capture rectified stereo images using a single mirror by placing the mirror parallel to the optics of the camera. We found that aligning the mirror by hand was adequate to obtain a depth map of the scene (see Fig. 18). To demonstrate that the stereo image is rectified, the correspondence search is only performed along the scan-lines. The depth map was computed on a 640×480 image using both SSD and normalized cross-correlation with a 15×15 size window. In the single mirror case, we found that normalized cross-correlation is beneficial because of intensity differences introduced by the reflection of scene rays at acute angles with the mirror (recall that the reflectance of a mirror falls slightly as a function of the angle of incidence).

To construct a three mirror system, we found the optimal configuration as described in the previous section, using a $70°$ field of view and a value of $c = 0.2$. We scaled the

system so that the baseline $b = 10$ cm and $c = 2$ cm. Using a drawing tool, we printed the location of the three mirrors and camera center of projection on a piece of paper in order to accurately place the mirrors. Fig. 17b shows the experimental setup.

Fig. 19a shows several images taken by the three mirror system. Figs. 19b and 19c show depth maps computed by searching along the scan-lines using both SSD and normalized cross-correlation. For the three mirror case, we found that normalizing the data did not improve the results. The three mirror system does not suffer from the acute angles of incidence which are encountered when using a single mirror. Therefore, the extra computational cost of normalized cross-correlation can be avoided.

8   CONCLUSION

In this paper, we have shown how to design a class of novel stereo sensors. By avoiding the need for synchronization, rectification, and normalization of the data, these sensors are well-suited for real-time applications. We have derived the constraints that must be satisfied in order to obtain rectified stereo images from a catadioptric system. In addition, we have shown how to use the remaining degrees of freedom to design compact stereo sensors. Through error analysis, we were able to show the feasibility of building these sensors.

REFERENCES

[1] 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.
[2] J. Gluckman and S.K. Nayar, "Planar Catadioptric Stereo: Geometry and Calibration," *Proc. 1999 Conf. Computer Vision and Pattern Recognition,* 1999.
[3] A. Goshtasby and W.A. Gruver, "Design of a Single-Lens Stereo Camera System," *Pattern Recognition,* vol. 26, no. 6, pp. 923–937, 1993.
[4] E. Hecht and A. Zajac, *Optics.* Addison-Wesley, 1974.
[5] M. Inaba, T. Hara, and H. Inoue, "A Stereo Viewer Based on a Single Camera with View-Control Mechanism," *Proc. Int'l Conf. Robots and Systems,* July 1993.
[6] F. Isgro and E. Trucco, "Projective Rectification Without Epipolar Geometry," *Proc. 1999 Conf. Computer Vision and Pattern Recognition,* 1999.
[7] D.H. Lee, I.S. Kweon, and R. Cipolla, "A Biprism Stereo Camera System," *Proc. 1999 Conf. Computer Vision and Pattern Recognition,* 1999.
[8] C. Loop and Z. Zhang, "Computing Rectifying Homographies for Stereo Vision," *Proc. 1999 Conf. Computer Vision and Pattern Recognition,* 1999.
[9] H. Mathieu and F. Devernay, "Systeme de Miroirs pour la Stereoscopie," Technical Report 0172, INRIA Sophia-Antipolis, 1995.
[10] H. Mitsumoto, S. Tamura, K Okazaki, N. Kajimi, and Y. Fukui, "3D Reconstruction Using Mirror Images Based on a Plane Symmetry Recovery Method," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 9, pp. 941–945, Sept. 1992.
[11] S.K. Nayar, *Robotic Vision System,* United States Patent 4,893,183, Aug. 1988.
[12] S.A. Nene and S.K. Nayar, "Stereo with Mirrors," *Proc. 6th Int'l Conf. Computer Vision,* Jan. 1998.
[13] M. Pollefeys, R. Koch, and L. VanGool, "A Simple and Efficient Rectification Method for General Motion," *IEEE Int'l Conf. Computer Vision,* pp. 496-501, 1999.
[14] L. Robert, M. Buffa, and M. Herbert, "Weakly-Calibrated Stereo Perception for Rover Navigation," *Proc. Fifth Int'l Conf. Computer Vision,* 1995.
[15] D. Southwell, A. Basu, M. Fiala, and J. Reyda, "Panoramic Stereo," *Proc. Int'l Conf. Pattern Recognition,* 1996.
[16] Z.Y. Zhang and H.T. Tsui, "3D Reconstruction from a Single View of an Object and Its Image in a Plane Mirror," *Proc. Int'l Conf. Pattern Recognition,* 1998.

Joshua Gluckman received the BA degree in economics from the University of Virginia, the MS degree in computer science from The College of William and Mary, and the PhD degree in computer science from Columbia University. He is currently a post-doctoral vision researcher at Columbia University. He is a member of the IEEE Computer Society.

Shree K. Nayar received the PhD degree in electrical and computer engineering from the Robotics Institute at Carnegie Mellon University in 1990. He is a professor in the Department of Computer Science at Columbia University. He currently heads the Columbia Automated Vision Environment (CAVE), which is dedicated to the development of advanced computer vision systems. His research is focused on three areas, namely, the creation of novel vision sensors, the design of physics-based models for vision, and the development of algorithms for scene understanding. His work is motivated by applications in the fields of digital imaging, computer graphics, human-machine interfaces, robotics, and image understanding. He has received the David Marr Prize twice (1990 and 1995), the David and Lucile Packard Fellowship (1992), the National Young Investigator Award (1993), and the Keck Foundation Award for Excellence in Teaching (1995). He is a senior member of the IEEE and a member of the IEEE Computer Society.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** http://computer.org/publications/dlib.