

# EPIPOLAR SPACES FOR ACTIVE BINOCULAR VISION SYSTEMS

*James Monaco, Alan C. Bovik*

University of Texas at Austin  
Department of Electrical  
and Computer Engineering  
Austin, TX 78712-1084 USA  
{monaco,bovik}@ece.utexas.edu

*Lawrence K. Cormack*

University of Texas at Austin  
Department of  
Psychology  
Austin, TX 78712-0187 USA  
cormack@mail.utexas.edu

## ABSTRACT

Depth recovery for active binocular vision systems is simplified if the camera geometry is known and corresponding points can be restricted to epipolar lines. Unfortunately, computation of epipolar lines requires calibration which can be complex and inaccurate. While it is possible to register images without geometric information, such unconstrained algorithms are usually time consuming and prone to error. In this paper we propose a compromise. Even without the instantaneous knowledge of the system geometry, we can restrict the region of correspondence by imposing limits on the possible range of configurations, and as a result, confine our search for matching points to *epipolar spaces*. For each point in one image, we define the corresponding epipolar space in the other image as the union of all associated epipolar lines over all possible system geometries. Epipolar spaces eliminate the need for calibration at the cost of an increased search region.

**Index Terms**— Stereo vision, Active vision, Image registration

## 1. INTRODUCTION

Active binocular vision systems (ABVS) provide a powerful means for extracting information from a complex scene. With a virtually unlimited field of view they have access to huge amounts of information, yet are able to confine their resources to specific regions of interest. Additionally, their ability to actively analyze their environment enables them to address, in a well-posed manner, tasks that may be ill-posed for a passive observer [1].

ABVS are especially well suited for the recovery of depth information. Depth recovery requires registering the two images, a notoriously difficult problem. This task can be greatly simplified if the camera geometry is known. With geometric knowledge of the stereo configuration, the search for corresponding points can be restricted to epipolar lines [2, 3]. Ascertaining the actively changing stereo geometry requires calibration. This calibration procedure can be highly complex,

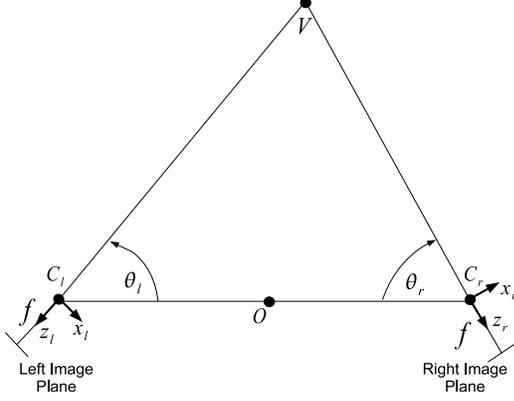
involving motorized lens calibration, kinematic calibration, and head/eye calibration [4]. Consequently, mechanical calibration requires highly accurate, often expensive equipment. For some ABVS, such as those deployed for planetary exploration using mobile rovers, calibration may not be feasible [5].

While it is possible to register images without calibration information, such unconstrained algorithms [6] are usually far more time consuming and prone to error. In this paper we propose a compromise. Even without knowledge of the exact stereo geometry we can restrict the region of correspondence by imposing limits on the possible range of configurations. That is, by restricting the range of vergence angles, baseline distances, and focal lengths etc., we can confine our search for matching points to what we refer to as *epipolar spaces*. For each point in one image, we define the corresponding epipolar space in the other image as the union of all associated epipolar lines over all possible system geometries.

## 2. EPIPOLAR SPACES

Knowledge of the camera geometry can be extremely valuable in the registration process, reducing the search for matching points to epipolar lines. Consider the simplified stereo geometry in Fig. 1. Here two pinhole cameras whose optical centers are located at  $C_l$  and  $C_r$  converge at the fixation point  $V$ . Both cameras have identical focal lengths  $f$ . The left and right camera rotation angles are  $\theta_l$  and  $\theta_r$ . Each camera has an associated right-hand coordinate system originating at its optical center. If  $\tilde{m}_l = [x_l, y_l, z_l, 1]^t$  and  $\tilde{m}_r = [x_r, y_r, z_r, 1]^t$  are projective world coordinates in the left and right camera frames, respectively, then their relation is  $\tilde{m}_r = [R|T]\tilde{m}_l$ , where

$$R = \begin{bmatrix} -\cos(\theta_l + \theta_r) & 0 & -\sin(\theta_l + \theta_r) \\ 0 & 1 & 0 \\ \sin(\theta_l + \theta_r) & 0 & -\cos(\theta_l + \theta_r) \end{bmatrix} \quad (1)$$



**Fig. 1.** Stereo geometry.

is the rotation matrix and

$$T = \begin{bmatrix} -\sin(\theta_r) \\ 0 \\ -\cos(\theta_r) \end{bmatrix} \quad (2)$$

is the translation vector. Let  $m_l = [u_l, v_l, 1]^t$  and  $m_r = [u_r, v_r, 1]^t$  be projective coordinates in the left and right image planes. It is well known that the equation relating corresponding epipolar lines is

$$m_r^t A_r^{-t} T_\times R A_l^{-1} m_l = 0, \quad (3)$$

where, in our simplified geometry,

$$A_l = A_r = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

are the identical intrinsic matrices and

$$T_\times = \begin{bmatrix} 0 & \cos(\theta_r) & 0 \\ -\cos(\theta_r) & 1 & \sin(\theta_r) \\ 0 & -\sin(\theta_r) & 0 \end{bmatrix} \quad (5)$$

implements the crossproduct as a matrix. The matrix  $E = T_\times R$  is the essential matrix and relates the coordinate frames. The matrix  $F = A_r^{-t} E A_l^{-1}$  is the fundamental matrix [7] and includes the intrinsic parameters of the cameras. The expression in (3) can be simplified to the following:

$$v_l = v_r \frac{f \sin(\theta_l) + u_l \cos(\theta_l)}{f \sin(\theta_r) - u_r \cos(\theta_r)}. \quad (6)$$

That is, for a given point  $(u_r, v_r)$  in the right image, the matching point in the left image (if not obscured) lies on the line given by (6).

We now consider the situation where the camera configuration actively changes and we no longer know the specific geometry. In such a situation we will not be able to restrict

our search for corresponding points to epipolar lines. However, even though we may not know the precise values of parameters such as focal length, baseline distance, and camera rotation angles, we can establish acceptable ranges for these values. Consequently, we can still restrict the location of corresponding points across images. For a given point in one image, the matching point in the other is confined to a region defined by the union of all corresponding epipolar lines produced over all possible camera configurations. We call these continuous regions *epipolar spaces*.

The goal of the remainder of this section is to quantify these epipolar spaces for a stereo rig with a fixed baseline and fixed focal length as shown in Fig. 1. In this configuration the only variable parameters that effect the epipolar geometry are the rotation angles  $\theta_l$  and  $\theta_r$ . Translation and rotation of the entire stereo rig about  $O$ , while allowed, do not influence the epipolar geometry. We establish the range of rotation angles by confining them to the interval

$$\theta_l, \theta_r \in [\theta_M, \pi - \theta_M], \quad (7)$$

where  $\theta_M$  is the minimum angle relative to the baseline.

Although theoretically a matching point can lie anywhere on the corresponding epipolar lines, the search is usually restricted to a maximum horizontal disparity. Bounding the horizontal disparity has the effect of limiting the depth around the horopter at which objects can be fused. For our purposes, we assume a maximum horizontal disparity defined by

$$|d| = |u_l - u_r| \leq D. \quad (8)$$

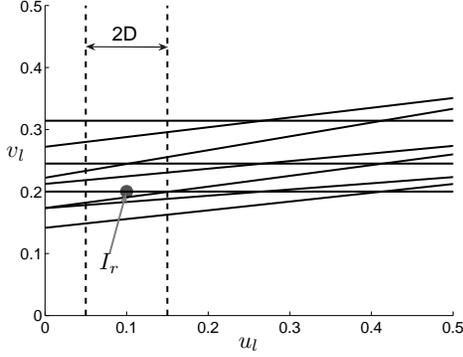
For the right image point  $I_r$  Fig. 2 illustrates several corresponding epipolar lines in the left image. Each separate epipolar line results from a unique geometric configuration, i.e. a unique combination of  $\theta_l$  and  $\theta_r$ . The dashed vertical lines delimit the maximum allowable horizontal disparities defined by (8).

The restriction imposed by (8) determines the leftmost and rightmost bounds of the epipolar spaces. The upper and lower bounds are determined by maximizing and minimizing (6) with respect to both  $\theta_l$  and  $\theta_r$ , respectively. The maximum can be found by separately maximizing the numerator and minimizing the denominator. Taking the derivative of the numerator with respect to  $\theta_l$ , setting it to zero, and solving for  $\theta_l$  yields the maximizing value  $\theta_l = \text{atan}(f/u_l)$ . The denominator (which must be positive so long as the cameras never image each other, i.e. the images do not contain the epipoles) is minimized when  $\theta_r = \theta_M$ . Inserting these results into (6), the upper bound becomes

$$v_{l,max} = v_r \frac{\sqrt{f^2 + u_l^2}}{f \sin(\theta_M) - u_r \cos(\theta_M)}. \quad (9)$$

The minimization of (6) is performed similarly with  $\theta_l = \pi - \theta_M$  and  $\theta_r = \text{atan}(-f/u_l)$ , producing

$$v_{l,min} = v_r \frac{f \sin(\theta_M) - u_r \cos(\theta_M)}{\sqrt{f^2 + u_l^2}}. \quad (10)$$



**Fig. 2.** Left image epipolar lines corresponding to right image point  $I_r$  for several unique geometric configurations. Specifically, the camera rotation angles  $\theta_l$  and  $\theta_r$  are each set to an element of  $\{\pi/4, \pi/3, \pi/2\}$ , producing nine combinations.

Since the horizontal extents of these bounds are limited by (8),  $f^2 + u_l^2$  in the preceding equations can be rewritten as  $f^2 + (u_r + \epsilon)^2$ , where  $\epsilon \in [-D, D]$ . In practice  $f \gg \epsilon$  and  $f \geq u_r$ , allowing the following approximation:  $f^2 + u_l^2 \approx f^2 + u_r^2$ . Incorporating this result into (9) and (10) produces

$$v_{l,max} \approx v_r \frac{\sqrt{f^2 + u_r^2}}{f \sin(\theta_M) - u_r \cos(\theta_M)} = v_r c(u_r) \quad (11)$$

$$v_{l,min} \approx v_r \frac{f \sin(\theta_M) - u_r \cos(\theta_M)}{\sqrt{f^2 + u_r^2}} = \frac{v_r}{c(u_r)}, \quad (12)$$

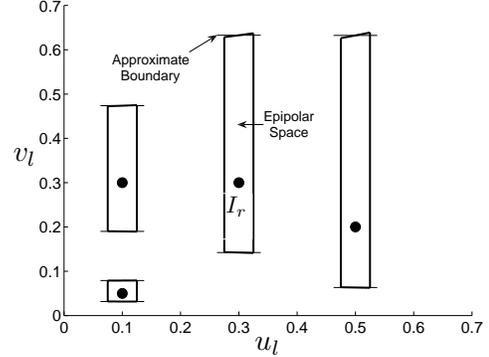
where

$$c(u_r) = \frac{\sqrt{f^2 + u_r^2}}{f \sin(\theta_M) - u_r \cos(\theta_M)}. \quad (13)$$

Remarkably, an epipolar space is well modeled by a rectangle. This fact is illustrated in Fig. 3. The thick lines represent the precise boundary of the regions described by (8), (9), and (10). The thin lines denote the approximate upper and lower bounds determined from (11) and (12). Epipolar spaces are nonuniform in area, increasing in size with increasing values of  $u$  and  $v$ .

### 3. CONCLUSION

In this paper we introduced the concept of an epipolar space. For a point in one image, the epipolar space was defined as the region in the other image formed from the union of all associated epipolar lines produced over all possible geometric configurations. Epipolar spaces eliminate the need for calibration, but at the expense of a greater search space.



**Fig. 3.** Various epipolar spaces. The large dots represent points from the right image. Each point is enclosed in its attendant epipolar space in the left image. The thick lines represent the precise boundaries described by (8), (9), and (10). The thin lines denote the approximate upper and lower bounds determined from (11) and (12).

### 4. REFERENCES

- [1] J. Aloimonos, I. Weiss, and A. Bandyopadhyay, "Active vision," *Proceedings of the First International Conference on Computer Vision*, pp. 35–54, 1987.
- [2] A. C. Bovik, *The Handbook of Image & Video Processing*. San Diego, CA: Academic Press, 2000.
- [3] B. Horn, *Robot Vision*. The MIT Press, 1986.
- [4] S.-W. Shih, Y.-P. Hung, and W.-S. Lin, "Calibration of an active binocular head," *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, vol. 28, no. 4, pp. 426–442, 1998. [Online]. Available: <http://dx.doi.org/10.1109/3468.686704>
- [5] Z. Zhang, Q.-T. Luong, and O. Faugeras, "Motion of an uncalibrated stereo rig: Self-calibration and metric reconstruction," *Proceedings of the International Conference on Pattern Recognition*, vol. 1, pp. 695–697, 1994. [Online]. Available: <http://dx.doi.org/10.1109/ICPR.1994.576407>
- [6] J. Magarey and A. Dick, "Multiresolution stereo image matching using complex wavelets," *Proceedings of the 14th International Conference on Pattern Recognition*, vol. 1, pp. 4–7, August 1998.
- [7] O. D. Faugeras, Q.-T. Luong, and S. J. Maybank, "Camera self calibration: Theory and experiments," *Proceedings of the Second European Conference on Computer Vision*, vol. 588, pp. 321–334, 1992.