

UNCALIBRATED STEREO RECTIFICATION FOR AUTOMATIC 3D SURVEILLANCE

Ser-Nam Lim[‡] Anurag Mittal[§] Larry S. Davis[‡] Nikos Paragios[§]

University of Maryland, College Park, CS Dept[‡] Siemens Corporate Research, Princeton, NJ[§]

ABSTRACT

We describe a stereo rectification method suitable for automatic 3D surveillance. We take advantage of the fact that in a typical urban scene, there is ordinarily a small number of dominant planes. Given two views of the scene, we align a dominant plane in one view with the other. Conjugate epipolar lines between the reference view and plane-aligned image become geometrically identical and can be added to the rectified image pair line by line. Selecting conjugate epipolar lines to cover the whole image is simplified since they are geometrically identical. In addition, the polarities of conjugate epipolar lines are automatically preserved by plane alignment, which simplifies stereo matching.

1. INTRODUCTION

Stereo rectification is a process that transforms a pair of stereo images such that conjugate epipolar lines (refer to [1] and Section 2) are aligned horizontally. It simplifies stereo matching. In automatic 3D outdoor surveillance, stereo correspondences are used for purposes such as 3D scene reconstruction (refer to [1, 2, 3]), background modeling that is invariant to local illumination changes, collaborative multi-camera control etc. Stereo rectification of image pairs thus becomes an important first step.

In practice, we need to minimize loss of pixel information and image distortion in the rectified images. In addition, our work is also motivated by the need for a rectification method that supports automatic 3D surveillance, in which cameras are dynamically positioned to optimize performance. We envision a configuration of moving cameras monitoring an outdoor scene such as those shown in Figure 1. The relative positions of two cameras chosen to accomplish certain 3D tasks might cause difficulties in the stereo correspondence phase. For example, the epipoles of two cameras can be on different side w.r.t. their respective image planes. As a result, corresponding pixels along conjugate epipolar lines beginning from the epipoles are in reverse order to each other i.e. the polarities of conjugate epipolar lines are reversed. This complicates intensity-based stereo matching. If the cameras are static, it is only necessary to build the stereo model once and therefore special steps can be taken to preserve the polarities of conjugate epipolar lines. However, if the cameras are moving, we desire a rectification method that will do so automatically to simplify subsequent computations.

Rectification is often achieved by transforming the images so that the epipoles become $[1, 0, 0]^T$, the point at infinity. In terms of epipolar geometry, the fundamental matrix for a rectified image pair becomes

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \quad ([2]).$$

One approach to stereo rectification is to apply a rectifying transformation to each image. To compute the transformations, they are often decomposed into simpler transforms. As described in [4], this often leads to some

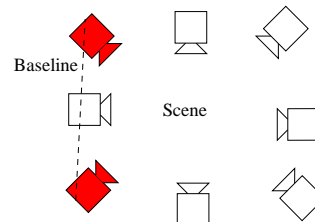


Fig. 1. A configuration of cameras. In this configuration, the red cameras have conjugate epipolar lines with corresponding pixels in reverse order since their epipoles are on different side of their respective images. A rectification method that preserve the polarities is desired.

nonlinear optimization problem. In practice, we found that such methods are often inaccurate and inefficient.

If the cameras are calibrated, the rectifying transformations can be computed by re-projecting a point in the scene onto the normal image planes as described in [5]. The cameras are defined to be in their respective normal image plane positions when they are arranged with their optical axes parallel and perpendicular to the baseline. Such approaches are often referred to as planar rectification ([6]). In the work described in [7], it was pointed out that planar rectification introduces significant image distortion as the forward motion component increases. This can potentially lead to unbounded rectified images. As a result, [7] uses a cylindrical rectification that is guaranteed to provide bounded rectified images and also ensures minimal loss of pixel information. This method is however relatively complex since all operations are performed in 3-D space.

Many stereo rectification methods fail when the epipoles are located in the images, as mentioned in [8], since this leads to infinitely large images. A direct method was introduced in [8] which can deal with all camera geometries. It described an approach where oriented half epipolar lines are added to the rectified images line by line based on a polar coordinate system around the epipoles. This is highly efficient and effective. Conjugate extremal epipolar lines have to be first determined because the epipoles are in different image positions. Here, we adopt the same direct method to create the rectified images by adding conjugate epipolar lines line by line. However, the method in [8] does not attempt to preserve the polarities of conjugate epipolar lines.

Our method exploits the fact that in a typical urban scene, there ordinarily exists a small number of dominant planes (e.g. ground, walls etc). Images of these planes in two views are related by 3×3 matrices, commonly known as homographies ([2]). By aligning a plane in a second view with the reference view using the corre-

sponding homography, the resulting warped image has geometrically identical epipolar lines as the reference view. This observation is also described in work on planar parallax such as [9, 10]. Consequently, conjugate epipolar lines are easily chosen to cover the whole image, unlike the method in [8] where common regions between extremal epipolar lines have to be first determined. The polarities of conjugate epipolar lines are preserved between the plane-aligned image and reference view, which provides an effective constraint for intensity-based stereo matching.

This paper is organized as follows. Section 2 provides a brief introduction to epipolar geometry and the effects of aligning planes on the polarities of conjugate epipolar lines. Section 3 then provides an introduction to planar parallax and its effect on conjugate epipolar lines between the reference view and plane-aligned image. Section 4 shows how conjugate epipolar lines are chosen to cover the whole image and how the starting position of each rectified image row can be computed to minimize image distortion. The conclusions are given in Section 5.

2. BACKGROUND

2.1. Epipolar Geometry

We will briefly describe epipolar geometry in this section. Details can be found in references such as [1, 2, 3]. Given a left and right view of a point P in the scene, let m and m' be the respective images as shown in Figure 2. The plane C_1PC_2 is known as the epipolar plane. Clearly, the intersection of C_1PC_2 with the right image plane is a line l' on which m' must lie. l' can be expressed as Fm , where F is a 3×3 matrix with rank 2, known as the fundamental matrix. Consequently, we have the following epipolar constraint

$$m'^T F m = 0. \quad (1)$$

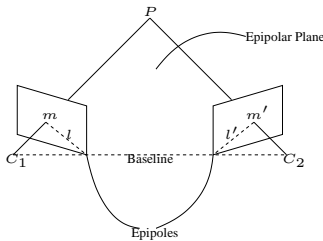


Fig. 2. Epipolar geometry: C_1 and C_2 are the camera centers and the line joining them is known as the baseline. The image positions where the baseline intersects the image planes are known as epipoles. The epipolar plane corresponds to C_1PC_2 .

F can be determined automatically with a minimum of 8 corresponding points by solving Equation 1 using the RANSAC algorithm described in [11]. Conversely, given F and m , we can determine l' . The conjugate l of l' can be determined similarly. Referring to Figure 2, we can see that the corresponding pixels of all pixels on l must lie on l' . As a result, the two views can be rectified by adding l and l' to the rectified image pair.

2.2. Plane Alignment

Given two views of a scene, there is a linear projective transformation H_Π (known as an homography) relating the projection m of the points on a plane Π in the first view to its projection in the second view, m' . This can be expressed as

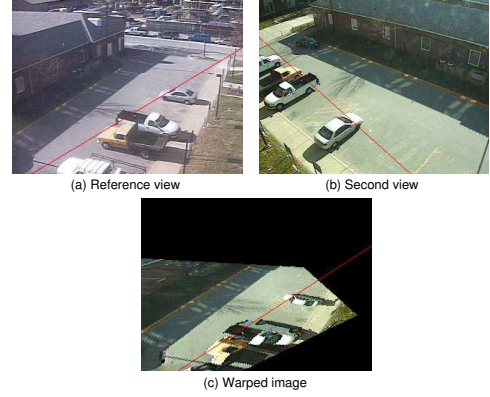


Fig. 3. Corresponding pixels on the conjugate epipolar lines (the red lines) between the reference and second view are in reversed order beginning from the epipoles (epipoles in the reference and second view are in octant 3 and 1 respectively, refer to Figure 5). Polarities of conjugate epipolar lines are however preserved between the reference view and the plane-aligned image.

$$H_\Pi \times m = m'. \quad (2)$$

H_Π can be determined automatically using RANSAC with a minimum of 4 corresponding images for points on Π . A plane-aligned image I can be derived from the second view using H_Π as follow

$$i_{aligned}(x, y, 1) = i_{second}(N(H_\Pi \times [x, y, 1]^T)), \quad (3)$$

where $(x, y, 1)$ is the homogeneous image position of a pixel in I and $i_{aligned}(x, y, 1)$ represents its intensity. i_{second} represents the intensity values in the second view and N represents the normalization of the image position after applying H_Π . Figure 3(c) shows an example of a plane-aligned image using the ground plane between the reference view in Figure 3(a) and the second view in Figure 3(b). It also show that plane alignment preserves the polarities of conjugate epipolar lines.

3. EFFECTS OF PLANAR PARALLAX

Given two views of a scene, we can compensate for the motion of a plane between them by applying the associated homography. The residual parallax displacement field in the plane-aligned image produced by Equation 3 is a displacement along a line geometrically identical to the epipolar line in the reference view. This translation is better known as an epipolar field. This is the basic principle of planar parallax and the proof ([9, 10]) is given as follows. Referring to Figure 4, let the plane-aligned image w.r.t. Π (using Equation 3) be I . Let p and q in the reference view be the images of P and Q , where Q lies on Π but not P . In I , the pixel at the image position corresponding to q is p' from the second view, since P is occluding Q . In the reference view, since p is the corresponding image of P , the residual displacement is equal to the distance pq along the epipolar line and is directly proportional to the distance of P from Π .

A direct consequence of this is that conjugate epipolar lines between the reference view and plane-aligned image (using Equation 3) are geometrically identical. An epipolar line in the plane-aligned image contains only two types of pixels: (1) pixels not

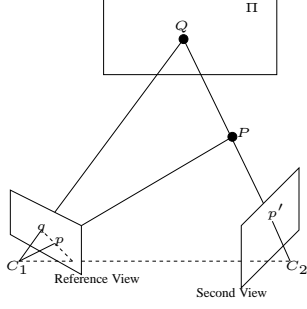


Fig. 4. Residual parallax field after plane alignment is epipolar. p and q are the images of P and Q in the reference view respectively, p' is the image of P in the second view and Π is the dominant plane.

lying on Π and (2) pixels on Π . For type (1) pixels, the corresponding pixel in the reference view lies on the same epipolar line plus some displacement. For type (2) pixels, the corresponding pixel in the reference view is in the same image position. It is also clear from Figure 4 that the displacements for type (1) pixels are always in a direction away from the epipoles¹.

4. CONSTRUCTING THE RECTIFIED IMAGES

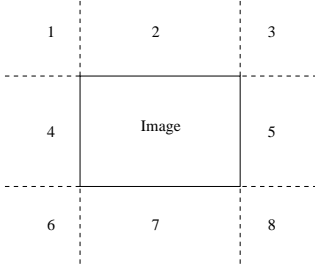


Fig. 5. Based on which octant the epipoles are in, selection of conjugate epipolar lines is straightforward. Octants 2, 4, 5 and 7 have two possible choices for the longest epipolar line. The longer one within the image boundaries is used. Exact positions of the epipoles are not explicitly needed since only the fundamental matrix is needed to determine the two epipolar lines.

Selecting conjugate epipolar lines We can compute the fundamental matrix F from the second view to the reference view as described in Section 2. Two points, m'_1 and m'_2 , in the second view are then chosen such that they have different epipolar lines in the reference view. These epipolar lines are given by Fm'_1 and Fm'_2 . The intersection of Fm'_1 and Fm'_2 gives the positions of the epipoles for both the reference view and plane-aligned image since conjugate epipolar lines between them are geometrically identical. In practise, it is often unreliable to determine the exact positions of the epipoles in this manner. However, it can reliably give us the octant (Figure 5) in which the epipoles lie. Once this

¹Formally, since the plane-aligned view is not a real view, the epipoles are undefined between the reference view and plane-aligned image.

is known, it becomes straightforward to select conjugate epipolar lines to cover the whole image. For example, if the epipoles are in octant 8, then we can select conjugate epipolar lines between the reference view and plane-aligned image corresponding to all pixels on the upper and left image boundary (note that choosing the left follow by the upper image boundary would turn the rectified image upside down). For each scanned pixel m , the corresponding conjugate epipolar lines are both $\hat{F}m$, where \hat{F} is the fundamental matrix from the reference view to the plane-aligned image. Since only \hat{F} is needed here, we do not explicitly need the positions of the epipoles. Rather, only the octant in which the epipoles lie is used to determine the image boundaries that need to be scanned. Each selected pair of conjugate epipolar lines is then added to their respective rectified images row by row.

Minimizing loss of pixel information Loss of pixel information can result from the scanning of each epipolar line. To prevent this, each epipolar line is scanned using Bresenham's line-scanning algorithm given in [12]. In addition, as mentioned in [8], the maximum distance between two consecutive selected epipolar lines must be ≤ 1 pixel (refer to Figure 6). While [8] resolved it by moving the epipolar line, we do so by choosing the proper image boundaries to scan. For example, if the epipole is in octant 8, then the upper and left image boundary should be scanned instead of the right and lower image boundary. On the upper boundary, the distance between two consecutively scanned pixels is 1 pixel. As a result, the distance between the other ends of the epipolar lines corresponding to those pixels is < 1 pixel since they are nearer to the epipoles. This guarantees that the maximum distance between the epipolar lines is ≤ 1 pixel. The same applies when the left boundary is scanned. Figure 6 illustrates this.

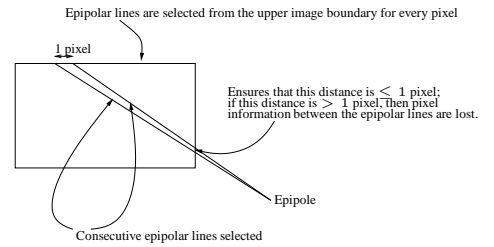


Fig. 6. Loss of pixel information occurs when the maximum distance between two consecutive epipolar lines are > 1 pixel. In our method, by choosing the correct image boundaries to scan, this is prevented.

Minimizing image distortion To minimize image distortion, we also need to determine the starting pixel position of each rectified image row. For illustration purpose, we will show an example where the lower boundary of the image is traversed pixel by pixel and the epipole is in octant 3 (refer to Figure 7). We first determine the length ℓ of the longest epipolar line L , which is that for the bottom-left pixel. The starting pixel position of L in the rectified image is the 0^{th} pixel position. Given an epipolar line L' of length ℓ' , we minimize image distortion by preserving the geometries between the rightmost pixels within the image boundary, label as a and b of L and L' respectively. This can be achieved as shown in Figure 7, giving the starting pixel position as $\ell - \ell' - \ell_{\perp}$ for the example, where ℓ_{\perp} is the distance between a and a point c such that the line segment bc is perpendicular to L . Note that ℓ and

ℓ' represent the lengths of their respective epipolar lines within the image boundary. Examples of our image rectification algorithm are given in Figure 5 and 9.

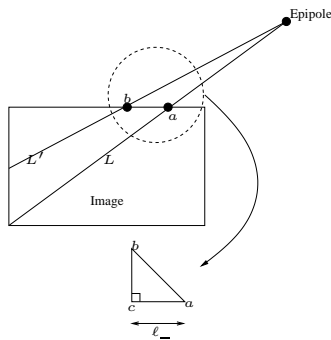


Fig. 7. Determine starting position of each rectified image row.

4.1. Epipoles in the Images

Our rectification method handles with ease when the epipoles lie in the corresponding images. In this case, we simply divide the image into two halves i_{upper} and i_{lower} along the horizontal line passing through the epipole. i_{upper} and i_{lower} can then be treated as image with epipole in octant 7 and 2 respectively. The rectified images can then be easily combined.

5. CONCLUSIONS

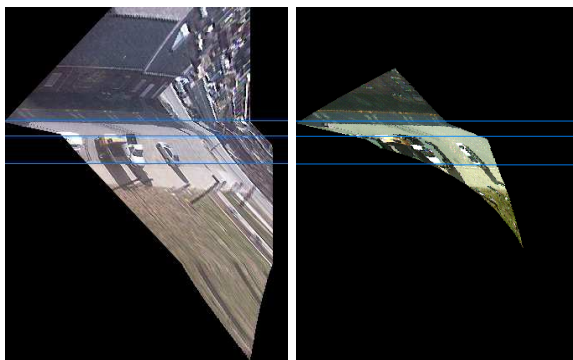


Fig. 8. The lines verified the correctness. The reference view, second view and plane-aligned image for (a) are the same as in Figure 3.

In conclusion, we have described a stereo rectification method that is efficient and effective for scenes with dominant planes. The main contribution lies in using a plane-aligned view, so that conjugate epipolar lines become geometrically identical and can be easily selected to cover the whole image while minimizing loss of pixel information. In addition, the polarities of conjugate epipolar lines are preserved automatically. As a result, our method is highly suitable for an automatic 3D surveillance system with multiple moving cameras.

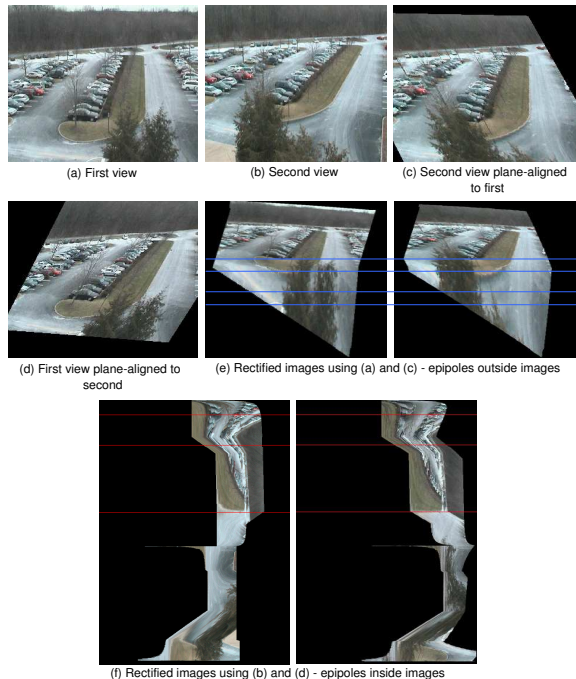


Fig. 9. When the second view is used as reference in the rectification process, the epipoles are located in the images. Our method handles it with ease. Note that the branches are different in both views and hence are occluded.

6. REFERENCES

- [1] Olivier Faugeras, *Three-Dimensional Computer Vision*, The MIT Press, Cambridge, Massachusetts, USA, 1993.
- [2] O. Faugeras, Q. T. Luong, and T. Papadopoulos, *The Geometry of Multiple Images*, MIT Press, 2001.
- [3] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521623049, 2000.
- [4] Charles Loop and Zhengyou Zhang, "Computing rectifying homographies for stereo vision," in *CVPR (1)*, 1999, p. 1125.
- [5] D. V. Papadimitriou and T. J. Dennis, "Epipolar line estimation and rectification for stereo image pairs," *IEEE Transactions on Image Processing*, vol. 5, no. 4, pp. 672–676, Apr 1996.
- [6] Ayache N. and Hansen C., "Rectification of images for binocular and trinocular stereovision," in *ICPR*, 1988, pp. 11–16.
- [7] S. Roy, J. Meunier, and I. Cox, "Cylindrical rectification to minimize epipolar distortion," in *CVPR*, 1997, pp. 393–399.
- [8] M. Pollefeys, R. Koch, and L. VanGool, "A simple and efficient rectification method for general motion," in *ICCV*, 1999, pp. 496–501.
- [9] R. Kumar, P. Anandan, and K. Hanna, "Direct recovery of shape from multiple views: A parallax based approach," in *ICPR*, 1994, pp. 685–688.
- [10] H. S. Sawhney, "Simplifying motion and structure analysis using planar parallax and image warping," in *Computer Vision and Pattern Recognition*, 1994, pp. 403–408.
- [11] Richard Hartley, "In defense of the 8-point algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 580–593, 1997.
- [12] Jack E. Bresenham, "Algorithm for computer control of a digital plotter," *IBM Systems Journal*, vol. 4, no. 1, pp. 25–30, 1965.