REFERENCES

[1] A. K. Jain, "Image data compression: a review," *Proc. IEEE,* Mar. 1981, vol. 69, no. 3, pp. 349-389.
[2] ——, "Fundamentals of digital image processing." London, U.K.: Prentice-Hall, 1989.
[3] S. L. Tanimoto, "Image transmission with gross information first," *Comput. Graphics Image Processing*, vol. 9, pp. 72–76, Jan. 1979.
[4] CCITT, "Standardization of group 3 facsimile equipment for document transmission," CCITT Recommendation T. 4.
[5] K. Liu and R. Prasad, "Comparing coding efficiency of vector chain coding and run-length coding for line drawings," *Proc. Inst. Elect. Eng.—I*, vol. 138, no. 5, pp. 363–370, Oct. 1991.
[6] H. Freeman, "On the encoding of arbitrary geometric configurations," *IRE Trans. Electron. Comput.*, vol. EC-10, pp. 260–268, June 1961.
[7] K. H. Tzou, "Progressive image transmission: a review and comparison of techniques," *Opt. Eng.*, vol. 26, no. 7, pp. 581–589, July 1987.
[8] C.-H. Chuang and C.-C. Jay Kuo, "Multiscale planar curve descriptor using biorthogonal wavelet transform," *Visual Inform. Processing II*, SPIE vol. 1961, pp. 44–54, 1993.
[9] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 7, pp. 675–69, July 1989.
[10] I. Daubechies, "Orthogonal bases of wavelets with finite support-connections with discrete filters," in *Proc. Int. Conf.: Wavelets: Time-Frequency Methods and Phase Space,* (J. M. Combes, A. Grossmann, Ph. Tchamitchian, Eds.,) Berlin, Germany: Springer, 1987.
[11] P. H. Westerink, "Subband coding of images," Ph.D dissertation, Delft Univ. Technology, Delft, The Netherlands, 1989.
[12] N. S. Jayant and P. Noll, *Digital Coding of Waveforms.* Englewood Cliffs, NJ: Prentice Hall, 1984.

Epipolar Line Estimation and Rectification for Stereo Image Pairs

D. V. Papadimitriou and T. J. Dennis

***Abstract*—The assumption that epipolar lines are parallel to image scan lines is made in many algorithms for stereo analysis. If valid, it enables the search for correspondent image features to be confined to one dimension and, hence, simplified. An algorithm that generates a vertically aligned stereo pair by warped resampling is described.**

I. INTRODUCTION

The centers of projection of the two cameras of a stereo viewing arrangement and any point in their field of view define an *epipolar plane*. The intersection of any such plane with the left or right image plane generates a line in 2-D image space called an *epipolar line* [1]. Thus, corresponding points of the two images belong to corresponding epipolar lines. This geometric relationship is called the *epipolar constraint*. Many stereo matching algorithms make the assumption that conjugate epipolar lines are collinear (or near-collinear), which enables them to restrict the search for homologous image points to one dimension [2]–[4]. The algorithm hence executes faster and, when the assumption is valid, gives more accurate results. When the application for the use of stereo is to compute absolute 3-D information about a scene, then camera calibration data must be available, and the use of the epipolar constraint relies on a simple geometrical computation. However, there are other applications [5], such as object recognition and photogrammetry, where accurate calibration is unnecessary or difficult to obtain and where the epipolar constraint has to rely on assumptions about the camera geometry. The present work aims to provide a solution for the latter case.

Manuscript received April 24, 1994; revised May 28, 1995. The associate editor coordinating the review of this paper and approving it for publication was Prof. John Juyang Weng.

The authors are with the Department of Electronic Systems Engineering, University of Essex, Colchester, U.K. (e-mail: tim@essex.ac.uk).

Publisher Item Identifier S 1057-7149(96)02764-9.

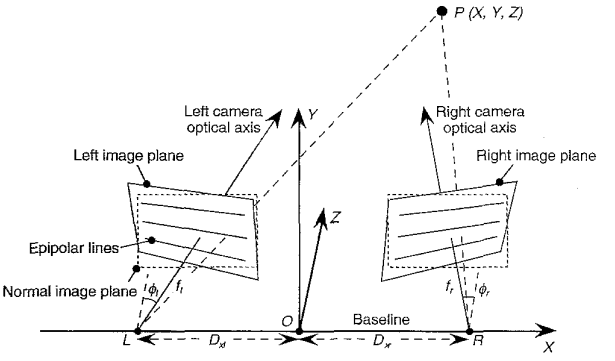

Fig. 1. Convergent stereo camera geometry. P represents a general world-point. With this camera arrangement the sets of epipolar lines are not parallel. Dashed lines represent the image planes after rectification.

As may be inferred from the above, stereo disparity analysis may be facilitated by first producing a vertically aligned stereo pair and then applying to it an algorithm that performs a 1-D search along the image scan lines in order to establish a dense field of point correspondences.

There is extensive literature on the reconstruction of a scene and/or the determination of object or camera motion from point correspondences between two views [6]–[9]. However, the task of establishing the transformations for rectification of the images is not trivial; therefore, our approach is limited to the determination of the continuously varying vertical misalignment in the commonly used convergent stereo camera system. This makes the problem more explicit but also simpler to approach since it avoids the use of strong assumptions about the imaging geometry. The advantages include simple implementation and fast rectification by means of image resampling.

In later sections, we describe a procedure that generates a vertically aligned stereo pair for uncalibrated camera arrangements where the assumption that conjugate epipolar lines are collinear is invalid.

II. EPIPOLAR GEOMETRY

Conjugate epipolar lines in the images of a stereo pair are collinear to each other (and consequently parallel to the horizontal scan lines of the images) only when the cameras are arranged with their optical axes parallel and perpendicular to the line connecting their optical centers *(baseline)* and are separated by a simple horizontal shift (a *parallel* stereo camera system). In any other case, epipolar lines are not collinear, which translates to a nonzero and position-dependent vertical coordinate of disparity. Making implicit use of the epipolar constraint by assuming that the vertical disparity component can be taken as zero can introduce severe errors during the matching process if the stereo camera arrangement is far from parallel.

Fig. 1 illustrates the camera arrangement for *convergent* stereo. The X axis of the world coordinate system is defined by the baseline. Throughout this section, it is assumed that rotation of either camera

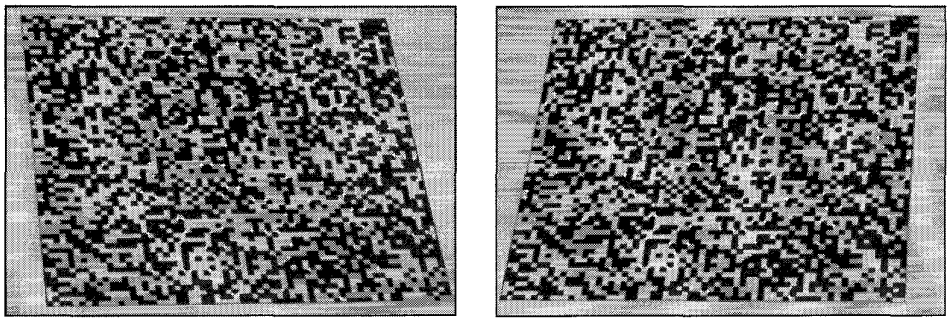

Fig. 2. Reduced-contrast original stereo pair of a synthetic scene of random squares with validated feature points superimposed. The left camera view is on the right, and vice versa, to enable proper optical fusion. Image size is 512 × 356.

occurs about an axis parallel to the world Y axis. Suppose the image planes indicated by solid lines are those corresponding to the true camera positions. Let (X, Y, Z) be the 3-D world coordinates of a point P, (x_{cl}, y_{cl}) and (x_{cr}, y_{cr}) its 2-D left and right camera coordinates, respectively, and let $(d_{cx}, d_{cy}) = (x_{cl} - x_{cr}, y_{cl} - y_{cr})$ define the disparity for point (x_{cl}, y_{cl}) in the left camera coordinate system. The situation is analyzed for the left camera; the right version is analogous. Assuming perspective projection and a pinhole camera model [10], from simple geometry, the following relationship is obtained:

$$\begin{bmatrix} X + D_{xl} \\ Y \\ Z \end{bmatrix} = \frac{s}{f_l} \cdot \begin{bmatrix} \cos\phi_l & 0 & -\sin\phi_l \\ 0 & 1 & 0 \\ \sin\phi_l & 0 & \cos\phi_l \end{bmatrix} \cdot \begin{bmatrix} x_{cl} - x_{clo} \\ y_{cl} - y_{clo} \\ f_l \end{bmatrix} \quad (1)$$

where

- D_{xl} distance (along the X axis) between the left camera optical center and the origin of the 3-D world coordinate system,
- s arbitrary scale factor,
- f_l effective focal length of the left camera,
- ϕ_l angle between the left camera optical axis and the Z axis of the world coordinate system, i.e., *pan* angle,
- (x_{clo}, y_{clo}) 2-D camera coordinates of the point where the left camera optical axis pierces the image plane, i.e., *image center*.

The arrangement where the cameras are parallel defines the *normal* image planes (indicated by dashed lines in Fig. 1). When the cameras are in this position, their optical axes are perpendicular to the baseline. Clockwise rotations from the normal positions are assumed negative; counterclockwise rotations are positive. By projecting point P onto the left normal image plane, we can calculate the new coordinates with respect to those in the original rotated image plane (see, for example, [11] for a detailed derivation of these equations)

$$x'_{cl} = f_l \cdot \frac{-f_l \cdot \sin\phi_l + (x_{cl} - x_{clo}) \cdot \cos\phi_l}{f_l \cdot \cos\phi_l + (x_{cl} - x_{clo}) \cdot \sin\phi_l} \quad (2)$$

$$y'_{cl} = \frac{f_l \cdot (y_{cl} - y_{clo})}{f_l \cdot \cos\phi_l + (x_{cl} - x_{clo}) \cdot \sin\phi_l} \quad (3)$$

where it is assumed that the origin of the new 2-D camera coordinate system coincides with the image center in the normal plane. In a similar manner, the equations of the rectified right camera coordinates can be derived.

When points (x_{cl}, y_{cl}) and (x_{cr}, y_{cr}) are projected onto the normal image planes, we expect them to lie on scan lines with the same vertical coordinate. Hence, if the camera parameters involved in (2) and (3) are known, then new vertically aligned images can be generated by resampling and interpolation of the original left and right images [12], [13]. These parameters can be obtained by a preliminary calibration stage, where a scene of known geometry is processed.

The approach adopted here is somewhat different. The target is to produce a stereo pair in which the images are vertically aligned without reference to any external data and in a noniterative process that precedes the stereo algorithm proper. This is accomplished by taking y'_l and y'_r to be equal

$$y'_{cl} = y'_{cr}$$

or, equivalently

$$\frac{f_l \cdot (y_{cl} - y_{clo})}{f_l \cdot \cos\phi_l + (x_{cl} - x_{clo}) \cdot \sin\phi_l} = \frac{f_r \cdot (y_{cr} - y_{cro})}{f_r \cdot \cos\phi_r + (x_{cr} - x_{cro}) \cdot \sin\phi_r} \quad (4)$$

After a few simple algebraic manipulations during which we replace y_{cr} with $y_{cl} - d_{cy}$ (only in those terms where y_{cr} appears as the only coordinate), we obtain an expression of the form

$$d_{cy} = a_0 + a_1 \cdot x_{cl} + a_2 \cdot y_{cl} + a_3 \cdot x_{cr} + a_4 \cdot x_{cl} \cdot y_{cr} + a_5 \cdot x_{cr} \cdot y_{cl} \quad (5)$$

where parameters $\{a_i,\ i = 0, \cdots, 5\}$ are functions of the focal lengths, the coordinates of the image centers, and the pan angles. Assuming that the 2-D camera coordinates (in length units) are related linearly to the 2-D image coordinates (in picture elements (pels)), an equation analogous to (5) holds if corresponding left and right image points are expressed in pel units ((x_l, y_l) and (x_r, y_r), respectively)

$$d_y = a_0 + a_1 \cdot x_l + a_2 \cdot y_l + a_3 \cdot x_r + a_4 \cdot x_l \cdot y_r + a_5 \cdot x_r \cdot y_l. \quad (6)$$

This equation enables determination of new (rectified) pel positions in the two images, given the corresponding positions in the original images and provided that parameters $\{a_i,\ i = 0, \cdots, 5\}$ are known.

Jones and Malik [14], Olsen [15], and Zhang *et al.* [16] have independently described analysis methods for the computation of the epipolar geometry. These methods are incorporated within stereo matching algorithms and use sparse [15], [16] or dense [14] point correspondences that have been established after an initial fully 2-D matching stage. In all of these approaches, once the epipolar geometry is estimated, it is used to recalculate disparities by restricting the search along the predicted epipolar lines. Hence, they are limited to determination of the epipolar line equation, and they differ from the present approach in both the analysis method and the extension to image rectification.

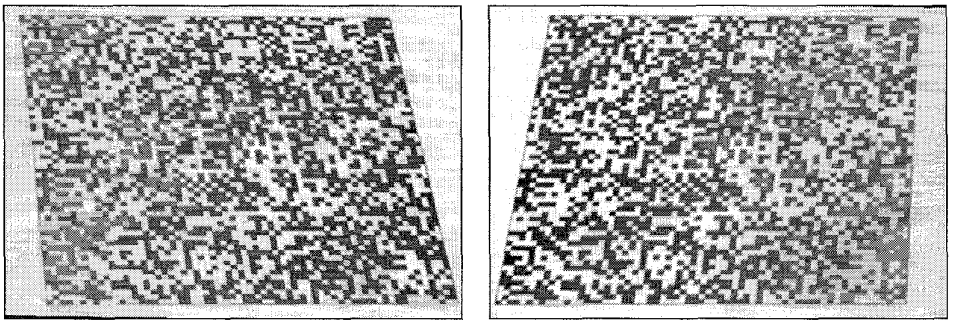

Fig. 3. Rectified stereo pair of synthetic scene.

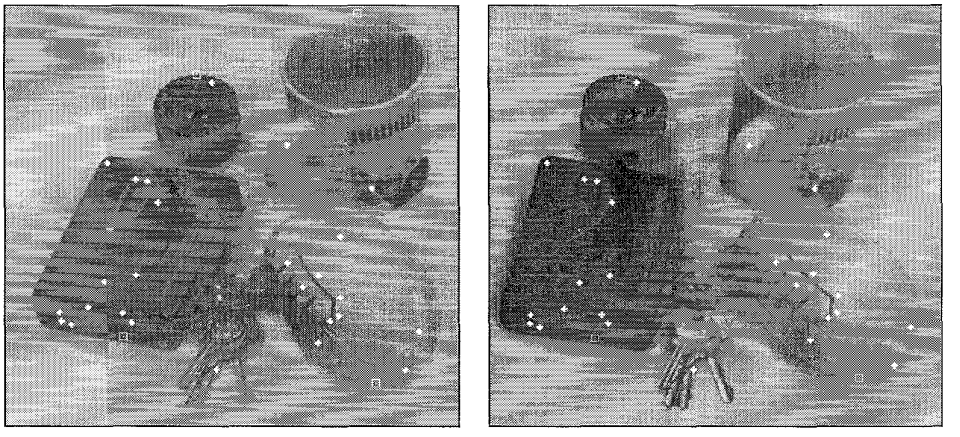

Fig. 4. Reduced-contrast original pair of an "objects" scene with validated feature points superimposed (crosses). The square symbols denote points where vertical disparity was measured manually to give the data in Table II. Image size is 512 × 480.

III. Vertical Registration Algorithm

Stage 1—Detection of Feature Points: The first step involves detection and localization of a number of unambiguous feature points (e.g., corners) in the images of the stereo pair. A slightly modified version of Förstner's interest operator [17] is currently used for identification of pels that correspond to image features. It has been preferred because its performance is well documented, it is easy to implement, and it has good sensitivity control. Moravec's interest operator [18] has also been tested, but its performance was inferior.

Stage 2—Matching: The feature points detected in stage 1 are used as input to a matching process applied between the two images. This is done twice, where each image takes its turn as reference. Every feature point in the reference image becomes the center of a subimage of size $(2K+1) \times (2K+1)$. Then, at the corresponding position in the second image, we perform a correlation operation between the reference subimage and all subimages within a search area of size $(2L_x+1) \times (2L_y+1)$. The *zero mean normalized cross correlation* (ZNCC) [19]—or simply the *correlation coefficient*—is employed as a similarity measure because of its useful properties:

- insensitivity to variations of the image intensities (dc-level and gain)
- value range restricted to $[-1, +1]$, which allows selection of a proper threshold.

A match is accepted only if the ZNCC function yields a value of 0.5 or more. Only matches that agree in both directions (*two-view matching consistency constraint* [20]) are accepted. In addition, matches are estimated to sub-pel accuracy by biquadratic interpolation of the local cross-correlation function. After this stage, we typically obtain a set of accurately matched and located, but sparsely distributed, image point pairs. The probability of a mismatch is low because of the severe constraints. Hence, the probability that spurious matches will exist among the final set of point pairs is low. This has been verified by experimentation.

TABLE I
Vertical Disparity Extrema (in PEL Units) with Respect to the Left Camera Image at the Corners of the Patterns in Fig. 2 (Column A) and the Marked Points in Fig. 4 (Column B)

	Random squares (A)	Objects (B)
Upper left	+5	0
Lower left	−8	−3
Upper right	−4	−4
Lower right	+3	+6

Stage 3—Epipolar Line Estimation: Point correspondences detected in stage 2 are used to calculate the vertical coordinate of disparity as a function of left and right image coordinates. The equation is of the form given in (6). A linear least squares procedure computes coefficients $\{a_i,\ i = 0, \cdots, 5\}$ using an LU-decomposition from [21]. The degrees of freedom for each camera are pan angle, focal length, and the image-plane coordinates of the principal point.

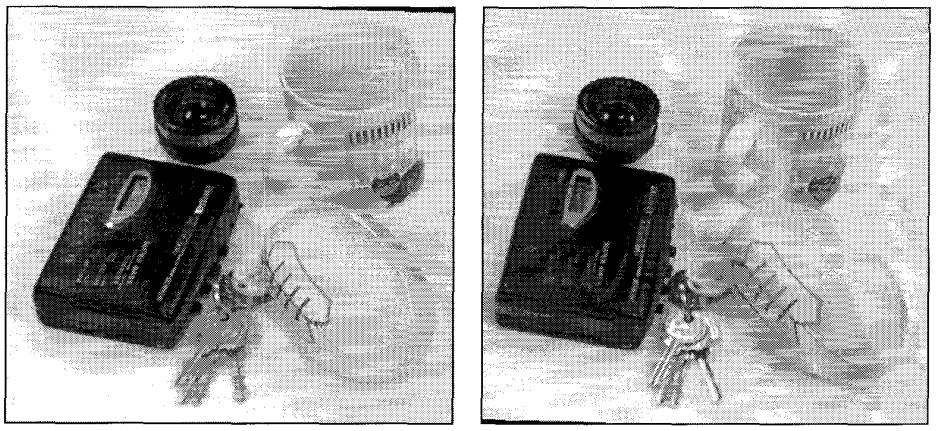

Fig. 5. Rectified pair of objects scene.

TABLE II
FEATURE POINT DETECTION AND VALIDATION STATISTICS FOR RANDOM SQUARES (COLUMN A) AND OBJECTS SCENES (COLUMN B)

	Random squares (A)	Objects (B)
Initially detected left feature points	987	644
Initially detected right feature points	934	680
Initially matched left feature points	633	524
Initially matched right feature points	632	524
Finally accepted matches	165	26
Mean error (picture elements)	0.1235	0.1607
Error standard deviation (picture elements)	0.1284	0.1677

Hence, eight point correspondences are the minimum required to provide a unique solution. However, there are usually many more.

Stage 4—Rectification of Stereo Pair: The model represented by the $\{a_i,\ i = 0, \cdots, 5\}$ guides a transformation process in which the two images of the stereo pair are resampled so that they become vertically aligned. Resampling of the left image uses the following equation:

$$\begin{pmatrix} x_l' \\ y_l' \end{pmatrix} = \begin{pmatrix} x_l \\ a_l + b_l x_l + c_l y_l + d_l x_l y_l \end{pmatrix} \tag{7}$$

and of the right image

$$\begin{pmatrix} x_r' \\ y_r' \end{pmatrix} = \begin{pmatrix} x_r \\ y_r + b_r x_r + d_r x_r y_r \end{pmatrix} \tag{8}$$

where $a_l = -a_0, b_l = -a_1 + a_0 a_4, c_l = 1 - a_2, d_l = -a_4$, and $b_r = a_3 + a_0 a_5, d_r = a_5$.

The following approximations have been made in order to derive the equations for image rectification

$$x_l y_r = x_l(y_l - d_y) \approx x_l(y_l - a_0) \tag{9}$$

$$x_r y_l = x_r(y_r + d_y) \approx x_r(y_r + a_0). \tag{10}$$

The underlying assumption is that a fixed vertical offset is a good first-order approximation of the vertical disparity. This is quite accurate since angles f_l and f_r are usually small. The approximations are necessary because rectification of the left (right) image requires that the new image coordinates are both given as a function of the original left (right) image coordinates *alone*. One may verify that (7) and (8) combined give a vertical disparity equation similar to (6), provided that the approximations of (9) and (10) are valid. Hence, the rectification process is applied independently in the left and right images without reference to any external parameters.

Resampling is performed by a 1-D linear interpolation procedure during which only two adjacent picture elements of one of the original images are involved in the calculation of the grey-level value of each pel of the new images. Ideally, the resampling process should be 2-D, where the operation in the horizontal direction should follow (2). However, this is not possible because the estimation process computes only six factors, whereas determination of eight camera parameters is required for full 2-D resampling.

IV. EXPERIMENTAL RESULTS

Two sets of stereo pairs were used to measure performance: one of a synthetic random squares pattern (Fig. 2) and the other containing man-made objects (Fig. 4). The pictures provide a demanding test for the process in the sense that the relative deformation in both horizontal and vertical directions between the two views is large. For the random squares scene, the camera arrangement was such that the angle between the optical axes of the cameras (the convergence angle $\phi_l + \phi_r$) was approximately 9.5°. The pattern-card was placed on the floor, with the camera mounting plate angled at 56° below the

horizontal. The corresponding angles for the objects scene were 7.8° and 45°. These combinations of convergence and oblique viewing angles gave rise to the vertical misalignment ranges (in pel units) listed in Table I. Although the convergence is large for a stereo viewing application, the angles involved are still small enough for the approximations in (9) and (10) to be valid.

The feature points accepted as correct after matching are overlaid on both sets of images. Table II shows the number of initially detected points, initially matched points (passed the correlation coefficient threshold), and finally accepted matches (passed the two-view matching consistency constraint). It also gives the mean and standard deviation of error. By "error," we mean the cumulative error that results both from the application of the least squares procedure as well as the approximations defined by (9) and (10). The equation of vertical disparity for the random squares was calculated to be

$$d_y = -0.810\,92 - 0.002\,38x_l - 0.010\,11y_l + 0.004\,92x_r + 0.000\,08x_l y_r + 0.000\,05x_r y_l.$$

For this derivation, the origin of each (left or right) image coordinate system has been assumed to be halfway along each image direction.

In Fig. 4, the feature points accepted as correct after matching are considerably fewer in number because image features are themselves quite sparse. The equation for the vertical disparity for the objects scene was calculated to be

$$d_y = 0.263\,88 - 0.009\,21x_l + 0.002\,73y_l + 0.012\,96x_r + 0.000\,06x_l y_r + 0.000\,06x_r y_l.$$

In the rectified stereo pairs, which are illustrated in Figs. 3 and 5, residual vertical misregistration (measured manually) is less than 1 pel everywhere.

V. Conclusion

An algorithm that generates a vertically aligned stereo pair has been described. In common with other known methods with similar goals, it is designed to work with images from camera systems with coplanar X and Z axes and parallel Y axes but that are otherwise uncalibrated. The approximations that have to be made mean that performance declines with increasing convergence angle but remains acceptable up to about 10°. However, 10° is an unusually "large" angle in this context.

The method uses grey scale image matching between the components of the stereo pair but confined to feature points. The matches are subjected to a rigorous cross checking procedure, which typically returns a relatively sparse but highly reliable set of corresponding coordinate pairs. Provided that the scene contains sufficient detail, the pairs will also be well-distributed spatially. The warping field is generated by a single-stage least-squares estimation in which all the confirmed pairs are involved. Because there is high confidence in their accuracy, there is no need for outlier detection and elimination.

The rectified stereo pair can be used as input to any disparity algorithm and allows it to exploit the epipolar constraint by the simplistic assumption of zero vertical coordinate of disparity. The algorithm has been tested with several stereo pairs, and the results have verified its robustness and generality.

References

[1] O. D. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*. Cambridge, MA: MIT Press, 1993.

[2] W. Hoff and N. Ahuja, "Surfaces from stereo: Integrating feature matching, disparity estimation and contour detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 2, pp. 121–136, 1989.

[3] M. S. Lew, T. S. Huang, and K. Wong, "Learning and feature selection in stereo matching," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 16, no. 9, pp. 869–881, 1994.

[4] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby, "PMF: A stereo correspondence algorithm using a disparity gradient limit," *Perception*, vol. 14, pp. 449–470, 1985.

[5] W. E. L. Grimson, "Why stereo is not always about 3-D reconstruction," A. I. Memo no. 1435, Artificial Intell. Lab., Mass. Inst. Technol., 1993.

[6] H. C. Longuet-Higgins, "A computer program for reconstructing a scene from two projections," *Nature*, vol. 293, pp. 133–135, 1981.

[7] R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of 3-D motion parameters of rigid bodies with curved surfaces," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-2, no. 6, pp. 13–27, 1984.

[8] J. Weng, T. S. Huang, and N. Ahuja, "Motion and structure from two perspective views: Algorithms, error analysis, and error estimation," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 11, no. 5, pp. 451–476, 1989.

[9] O. D. Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig?," in *Proc. Second Euro. Conf. Comput. Vision*, G. Sandini, Ed., Lecture Notes in Computer Science, Springer-Verlag, Santa Margherita Ligure, Italy, May 19–22, 1992, vol. 588, pp. 563–578.

[10] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[11] A. Goshtasby and W. B. Gruver, "Design of a single-lens stereo camera system," *Patt. Recogn.*, vol. 26, no. 6, pp. 923–937, 1993.

[12] L. Yan and Z. Zuxun, "Fast implementation for generating epipolar line images with one-dimensional resampling," in *Proc. Sixteenth Cong. Int. Soc. Photogrammetry Remote Sensing*, Kyoto, Japan, 1988, pp. 511–520.

[13] N. Ayache and C. Hansen, "Rectification of images for binocular and trinocular stereo vision," in *Proc. Ninth Int. Conf. Patt. Recogn.*, Rome, Italy, Nov. 17–19, 1988, pp. 11–16.

[14] D. G. Jones, "Computational models of binocular vision," Ph.D. Thesis, Stanford Univ., Stanford, CA, 1991, ch. 3.

[15] S. I. Olsen, "Epipolar line estimation," in *Proc. Second Euro. Conf. Comput. Vision*, G. Sandini, Ed., Lecture Notes in Computer Science, Springer-Verlag, Santa Margherita Ligure, Italy, May 19–22, 1992, vol. 588, pp. 307–311.

[16] Z. Zhang, R. Deriche, O. Faugeras, and Q. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," Res. Rep. 2273 (Robotics, Image, and Vision Program), INRIA, Sophia Antipolis, France, 1994.

[17] W. Förstner and E. Gulch, "A fast operator for detection and precise location of distinct points, corners and centres of circular features," in *Proc. Int. Soc. Photogrammetry Remote Sensing Intercommission Workshop*, Interlaken, Switzerland, June 1987, pp. 281–304.

[18] H. Moravec, "Toward automatic visual obstacle avoidance," in *Proc. Fifth Int. Joint Conf. Artificial Intelligence*, Cambridge, MA, Aug. 22–25, 1977, p. 584.

[19] D. H. Ballard and C. M. Brown, *Computer Vision*. Englewood Cliffs, NJ: Prentice-Hall, 1982.

[20] M. J. Hannah, "A system for digital stereo image matching," *Photogrammetric Engineering and Remote Sensing*, vol. 55, no. 12, pp. 1765–1770, 1989.

[21] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Programming*, 2nd ed. Cambridge, UK: Cambridge University Press, 1992.