REFERENCES

[1] P. D. Hyde and L. S. Davis, "Subpixel edge estimation," *Pattern Recognition,* vol. 16, no. 4, pp. 413–420, 1983.

[2] I. Overington and P. Greenway, "Practical first-difference edge detection with subpixel accuracy," *Image Vision Comput.,* vol. 5, no. 3, pp. 217–224, 1987.

[3] C. A. Berenstein, L. N. Kanal, D. Lavine, and E. C. Olson, "A geometric approach to subpixel registration accuracy," *Comput. Vision, Graphics, Image Processing,* vol. 40, no. 3, pp. 334–360, 1987.

[4] Q. Tian and M. N. Huhns, "Algorithms for subpixel registration," *Comput. Vision, Graphics, Image Processing,* vol. 35, no. 2, pp. 220–233, 1986.

[5] J. W. Hill, "Dimensional measurements from quantized images," in *Machine Intelligence Research Applied to Industrial Automation (SRI 10th Report for NSF Grant DAR78-27128),* D. Nitzan *et al.,* 1980, pp. 75–105.

[6] H. S. Ranganath, "Hardware implementation of image registration algorithms," *Image Vision Comput.,* vol. 4, no. 3, pp. 151–158, 1986.

[7] C. Ho, "Precision of vision system," in *1982 Workshop Industrial Applications of Machine Vision Conf. Rec.,* IEEE Cat. No. 82CH1755-8, pp. 153–159.

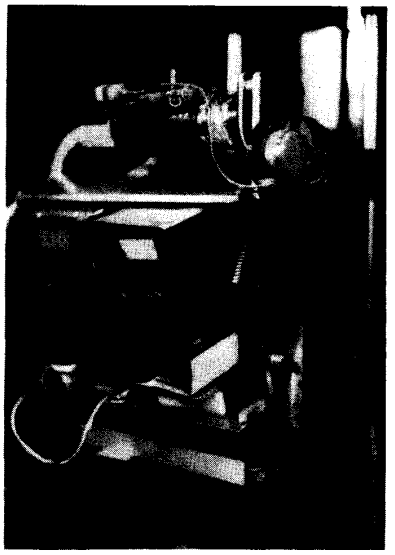


Fig. 1. The agile camera system. The agile camera system allows two cameras to translate horizontally and vertically, and to rotate by panning and tilting. Motors mounted in the camera housings allow control of the lenses by independently adjusting their focusing distance, focal length (zoom), and aperture diameter.

Stereo Ranging with Verging Cameras

Eric Krotkov, Knud Henriksen, and Ralf Kories

Abstract—**We present a novel method to compute absolute range from stereo disparities with verging cameras. The approach differs from others by concentrating, through both analysis and experiment, on the effects caused by convergence, rather than on the general camera calibration problem. To compute stereo disparities we first extract linear image features and then match them using a hypothesize-and-verify method. To compute range we derive the relationship between object distance, vergence angle, and disparity. Experimental results show the precision of the range computation, excluding mistaken matches, to be approximately 5% for object distances up to three meters and a baseline distance of 13 cm. Including mistaken matches results in performance an order of magnitude worse, leading us to suggest methods to identify and model them.**

Index Terms—**Binocular stereo, stereo matching, stereo reconstruction, three-dimensional computer vision, vergence.**

I. INTRODUCTION

Computing range from stereo requires first matching the images taken by two cameras to determine disparities (differences in the positions of corresponding features), and then transforming these into absolute distances. A great deal of research in computer vision, robotics, photogrammetry, psychology, and neurophysiology has addressed both of these problems. Our contribution to this research is to develop and analyze a novel method for absolute stereo ranging with a pair of verging cameras.

Manuscript received December 10, 1986; revised April 19, 1990. Recommended for acceptance by O. D. Faugeras. This work was performed in the Grasp Laboratory at the University of Pennsylvania and was supported in part by NSF/DCR, U.A. Air Force, DARPA/ONR, U.S. Army, NSF/CER, DEC Corp., IBM Corp., Lord Corp., and the German Federal Ministry of Defense.

E. Krotkov is with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213.

K. Henriksen is with the Department of Computer Science, University of Copenhagen, Universitetsparken 1, DK-2100 Copenhagen, Denmark.

R. Kories is with the Deutsche Bundespost Telekom, D-6110 Dieburg, Germany.

IEEE Log Number 9040026.

Fig. 1 illustrates the physical camera system, which is described in detail elsewhere [9]. Two cameras mounted on a platform can translate horizontally and vertically, and rotate left-right and up-down. Motors mounted on each lens adjust the focal length, focusing distance, and aperture diameter. Further, the two cameras can *verge,* by rotating towards each other (converging) or away from each other (diverging). Fig. 2 illustrates the vergence mechanism; the travel from minimum to maximum vergence angle is approximately 6°, covering 90 000 motor steps. Potential advantages of vergence include increasing the field of view common to two cameras, and constraining the stereo disparity.

We model each lens as a pinhole, assuming that to first order all lines of sight intersect at a unique lens center. The lens centers are separated by a baseline distance b, and both lenses have focal length f. Associated with the cameras are reference frames L and R, with origins at the lens centers, and Z-axes coincident with the optic axes, positive in the direction of the scene. We define a Cyclopean reference frame W with origin at the midpoint of the baseline, Z-axis normal to the baseline, and X-axis coincident with the baseline (Fig. 3).

We address the following problem: identify the three-dimensional position of an object point $\boldsymbol{P} = [X_W, Y_W, Z_W]^T$ in the Cyclopean frame from its projections (x_L, y_L) and (x_R, y_R) onto the left and right image planes, respectively, using verging cameras. In addition, we are interested in the uncertainty on the measurement of Z_W.

This correspondence presents a novel method to compute Z_W that differs from classical approaches to stereo ranging that involve solving the camera calibration problem [5], [11]–[13] or related orientation problems [7, p. 311]. The approach differs by concentrating, through both analysis and extensive experiments, on the effects caused by convergence.

The correspondence is organized as follows. In Section II we describe a method for computing stereo disparities based on extracting and matching lines. We present in Section III the method for computing range as a function of disparity and vergence angle, and describe in Section IV a procedure to identify the parameters required for the

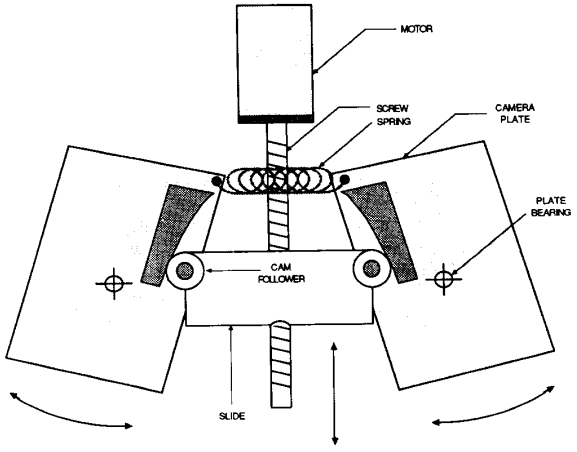


Fig. 2. Vergence mechanism. Top view of the vergence mechanism, which allows the two cameras to converge and diverge. The cameras are secured to the camera plates so that their optic axes point toward the top of the page. The D.C. servomotor turns the screw, which displaces the slide, causing the two camera plates (and thus the cameras) to rotate about the plate bearings.

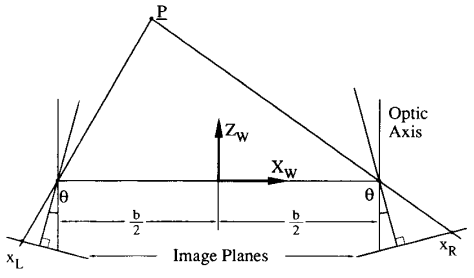


Fig. 3. Reference frames.

range computation. We provide quantitative experimental results in Section V, and conclude by summarizing the method and results, and discussing directions for future research.

II. Computing Stereo Disparities

To match two images, we must identify features in the two images that are projections of the same entity in the three-dimensional world. Barnard and Fischler [2] and Dhond and Aggarwal [4] survey numerous possible solutions. This section briefly describes one method for computing stereo disparities based on matching lines, which is detailed by Henriksen [6], and presents an example of its operation. The justification for using lines is that they are more likely to produce reliable matches than are points.

There are two basic steps in this approach: line extraction and line matching. The line extraction procedure (similar to [3]) describes the image by a set of line segments. The matcher proceeds by recursive prediction and verification (similar to [1]): first, it generates hypothetical matches that satisfy geometric constraints on the epipolar lines, and the differences in line segment length, orientation, and gradient magnitude; second, it prunes the hypotheses that do not satisfy local constraints of uniqueness and continuity; and third, it resolves ambiguities using global constraints.

The final outcome of the stereo matching process is a list of matching line segments and their disparities. The disparity is the distance in the image plane between two corresponding features. For the purpose of this correspondence, the disparity of two line segments is computed as the horizontal component of distance between their midpoints. This is a simplification, since the midpoints of corresponding line segments are seldom the images of the same physical point, and because the vertical disparity component is ignored.

Fig. 4 illustrates an example of the line matcher's performance. Figs. 4(a) and (b) show two images digitized to 256 gray-levels, with no enhancement or preprocessing. Figs. 4(c) and (d) show the line segments extracted from the images, 260 from each. On a VAX 11-750 without floating point accelerator this took 619.9 seconds. The line matcher determines the correspondence between these segments over a maximum disparity of 170 pixels, with maximum allowed difference in length of 10 pixels, maximum allowed angular deviation of 10°, and maximum allowed gradient magnitude difference of 5 units. The matching took 409.5 seconds for this example. Figs. 4(e) and (f) illustrate the 141 left–right and 144 right–left matches. The intersection of the two sets of matches contains three mistaken matches.

We believe that this example illustrates typical performance of the matcher. Qualitatively, we can describe its performance as fairly good at finding lines, and very conservative in matching lines. Naturally, prudent tuning of the extraction and verification parameters can improve performance significantly on any single stereo pair. To the extent that automatic adaptive tuning is not yet possible, we employ in all cases the same values as in this example for the experiments described below.

III. Computing Range

In this section we derive the analytical relationship between range, disparity, and vergence angle. We then formulate a procedure to identify from available measurements the parameters required for the range computation.

A. Derivation of Range Equation

Suppose that we rotate the right camera about the right lens center by θ, and that we rotate the left camera about the left lens center by $-\theta$ (Fig. 3). Then the coordinates of the point $\boldsymbol{P} = (X_W, Y_W, Z_W)^T$ expressed in the left and right camera coordinate systems, respectively, are

$$\begin{bmatrix} X_L \\ Y_L \\ Z_L \end{bmatrix} = \begin{bmatrix} \left(X_W + \frac{b}{2}\right)\cos\theta - Z_W \sin\theta \\ Y_W \\ \left(X_W + \frac{b}{2}\right)\sin\theta + Z_W \cos\theta \end{bmatrix},$$

$$\begin{bmatrix} X_R \\ Y_R \\ Z_R \end{bmatrix} = \begin{bmatrix} \left(X_W - \frac{b}{2}\right)\cos\theta + Z_W \sin\theta \\ Y_W \\ -\left(X_W - \frac{b}{2}\right)\sin\theta + Z_W \cos\theta \end{bmatrix}.$$

Perspective transformation yields the following image coordinates:

$$x_L = f\frac{\left(X_W + \frac{b}{2}\right)\cos\theta - Z_W \sin\theta}{\left(X_W + \frac{b}{2}\right)\sin\theta + Z_W \cos\theta},$$

$$x_R = f\frac{\left(X_W - \frac{b}{2}\right)\cos\theta + Z_W \sin\theta}{-\left(X_W - \frac{b}{2}\right)\sin\theta + Z_W \cos\theta}.$$

We solve both equations for X_W and equate them, thus eliminating X_W. Then we solve for Z_W:

$$Z_W = \frac{b}{\frac{f\sin\theta - x_R\cos\theta}{f\cos\theta + x_R\sin\theta} + \frac{f\sin\theta + x_L\cos\theta}{f\cos\theta - x_L\sin\theta}}$$

$$= \frac{b(f\cos\theta - x_L\sin\theta)(f\cos\theta + x_R\sin\theta)}{f(x_L - x_R)\cos 2\theta + (f^2 + x_L x_R)\sin 2\theta}. \quad (1)$$

This is the distance, normal to the baseline, from the baseline to the object point. It is convenient to measure distance not from the baseline, but from a plane π attached to the platform that supports the cameras and parallel to the baseline. We accomplish this by writing $Z = Z_W + Z_0$, where Z_0 is the *baseline offset* (distance from the baseline to π).

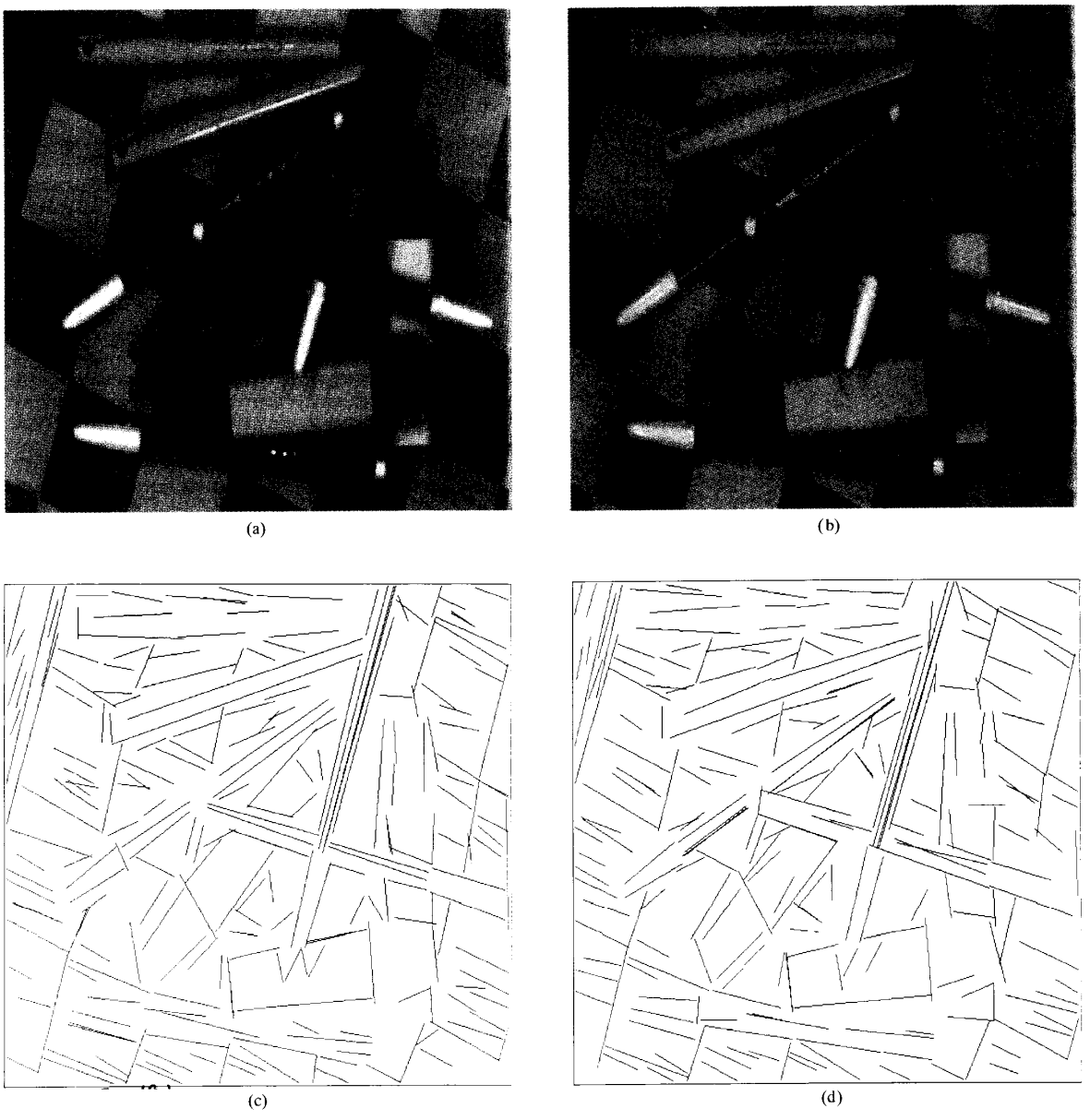


Fig. 4. Example of matching line segments. (a) Left image. (b) Right image. (c) Line segments extracted from the left image. (d) Line segments extracted from the right image. (e) Left–right matches (141) shown in the left image. (f) Right–left matches (144) shown in the right image. A postprocessor considers the set intersection of the matched segments in (e) and (f), and computes the disparity of two corresponding line segments as the horizontal component of distance between their midpoints.

B. Relationship of Parameters to Observations

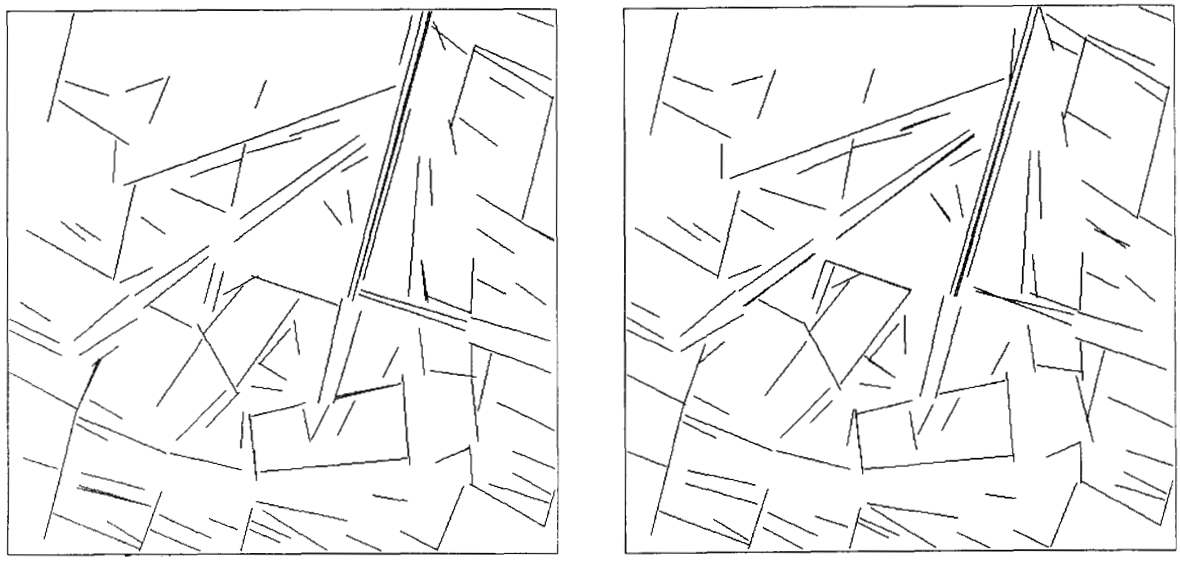
We summarize the derivation just completed as a function ξ:

$$Z = Z(Z_W, Z_0) = \xi(x_L, x_R, \theta, f, b, Z_0). \tag{2}$$

We do not observe directly any of the parameters of ξ. We infer x_L and x_R from image coordinates, we infer f and θ from motor positions, and we do not know b or Z_0. In this section we rewrite ξ in terms of parameters that we observe directly, and develop constraints on the parameters of ξ. We consider the parameters in the order in which they appear in (2).

We express the feature coordinates x_L and x_R, measured in mm, in a lens-centered frame (Fig. 3). The matcher finds corresponding points (u_L, v_L), and (u_R, v_R), measured in pixels, in an image-centered reference frame. Appendix A derives the transformation from image coordinates (pixels) to camera coordinates (mm).

We measure directly the vergence motor position V, not the vergence angle θ. The shape of the bearing guides on the camera plates (Fig. 2) determines the relationship between V and θ. Previous work [10, Appendix 2.4] shows that the relationship is quadratic, $\theta(V) = k_1V^2 + k_2V + k_3$, and identifies the values of k_i.



(e) (f)

Fig. 4. (*Continued.*)

We measure directly the zoom motor position F, not the focal length f. Previous work [10, Appendix 2.2] discusses the relationship between F and f. For the work reported here we set f to 20.0 (mm), and do not servo further the zoom motor.

The baseline is the line between the two lens centers; the baseline distance is b. The lens center is not a well-defined, constant, physical location; it actually moves as the lens focuses and zooms, making it difficult physically to measure b. In contrast, the camera rotation center *is* a well-defined location, about which the vergence mechanism rotates the cameras to converge them. We measure a distance r of 128 mm between the rotation centers, and estimate that the distance s normal to the optic axis between the lens center and the rotation center can be no greater than ± 5 mm. Thus, for parallel camera stations $r - 2s \leq b \leq r + 2s$ (mm). Appendix B shows that this constraint suffices for our camera system, i.e., in this case we can neglect constraints on b that arise uniquely from nonparallel camera stations.

Empirical observation suggests that the platform plane π is approximately 100 mm behind the baseline. So we constrain Z_0 by $-200 \leq Z_0 \leq 0$ (mm).

We now can write the range Z as a function η,

$$Z = \eta(u_L, u_R, k_1, k_2, k_3, V, f, b, Z_0), \tag{3}$$

where we observe u_L and u_R from images, we know k_1, k_2, k_3 and f from previous work, we observe V from the vergence motor encoder, and we do not know b or Z_0.

IV. Parameter Identification

This section presents a least-squares method to identify the unknown parameters b and Z_0. If we did not know k_1, k_2, k_3, or f, then we could use the same method to identify them.

First, we acquire ground-truth data. We place M planar objects in the field of view common to the cameras, and orient each visible face to be parallel to π (the plane attached to the platform and parallel to the baseline, cf. Section III-A). Let this be Scene A (Fig. 5). Then we manually measure the distance $Z_j^*, 1 \leq j \leq M$, normal to π to the jth object.

Second, we acquire disparity data for this scene viewed with different vergence angles. We servo the vergence motor to N different positions. At each, we digitize a stereo pair of images, and identify conjugate image points for each of the M objects. We supervise the selection of conjugate points to guarantee no mistakes in solving the correspondence problem and to bound tightly the feature localization error. The outcome of this procedure is $M \times N$ conjugate points $(U_L, v_L), (u_R, v_R)$ and the associated vergence motor positions V.

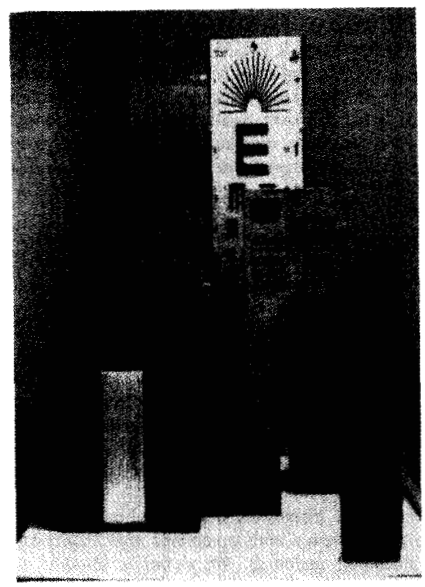


Fig. 5. Image of Scene A. The objects possess linear features at various orientations, and one possesses circular features (a challenge for the line finder).

Third, we search for the least-squares values for b and Z_0. Define the error at the ijth data point as $e_{ij} = Z_j^* - Z_{ij}$, where Z_{ij} is the distance computed by (3) to the jth object point while at the

ith vergence position, and Z_j^* is the manually measured distance. We employ a constrained minimization procedure (routine ZXMWD from the IMSL package) to identify the values of b and Z_0 that minimize E_{sse} within the bounds on b and Z_0 defined in Section III-B:

$$E_{sse} = \sum_{i=1}^{N} \sum_{j=1}^{M} e_{ij}^2, \tag{4}$$

where N is the number of vergence positions and M is the number of object points.

V. Experimental Results

In this section we present experimental results for the range computation using disparities computed with supervision, with no supervision, and with partial supervision. We present results for Scene A and Scene B, which contains the same objects as Scene A, but differently arranged.

We quantify the results in terms of the distributions of absolute errors and relative errors. The absolute error is $e = Z^* - Z$, where Z^* is the manually measured distance, and Z is the distance computed by (3). The relative error is $r = 100e/Z^*$ (percent). We compute the sample mean μ_E and sample standard deviation σ_E of the distribution of absolute errors; similarly, we compute μ_R and σ_R for the distribution of relative errors. The mean represents the accuracy of the range computation, and the standard deviation expresses its precision.

First, we compute disparities with supervision of both feature extraction and matching, as in Section IV. This guarantees no mistakes in solving the correspondence problem. For Scene A, $M = 7$ and $N = 9$; for Scene B, $M = 8$ and $N = 8$. Fig. 6 graphs the errors for Scene A, and Table I shows the statistics for the error distributions for both scenes. It shows that the range computation is highly accurate (almost 0% difference between known and measured ranges), which suggests that it does not introduce systematic error. It also shows that the precision of the range computation is about 2%, which quantifies the effect of imperfect parameter identification, imperfect in the sense that $E_{sse} \neq 0$ in (4).

Second, we compute disparities with no supervision whatsoever, running the matcher with default parameters (the same for all images) that enforce the constraints mentioned in Section II. Table II shows the statistics for the error distributions. Both accuracy and precision are significantly poorer than with supervision, particularly for Scene A. This decrease in precision is due to the presence of false matches, or *mistakes,* which result in unreasonable measurements (outliers). For example, we compute the position of one point to be across the street from our laboratory. The large standard deviations for Scene A show that mistaken matches can cause large range errors; the smaller standard deviations for Scene B show that this is not necessarily the case.

Third, we compute disparities with partial supervision, i.e., unsupervised feature extraction with supervised matching (postprocessing to remove mistaken matches). We define a mistake as a match resulting in $\sigma_R > 25\%$. Using this definition, we find that 5.1% of the matches are mistaken for Scene A, and 5.4% for Scene B. Table III shows the statistics for the error distributions after removing disparities due to mistaken matches. Both accuracy and precision are significantly better than without supervision; the precision is on the order of 5% for both scenes.

In summary, the error of the range computation with supervised feature extraction and supervised matching is approximately two percent; this includes errors due to imperfect parameter identification. The error with unsupervised feature extraction and supervised matching is approximately 5%; this includes errors due to imperfect feature localization and imperfect parameter identification, but excludes mistaken matches. The error with no supervision can be substantially larger, and may vary significantly across scenes. For this type of scene, approximately 5% of all matches are mistakes.

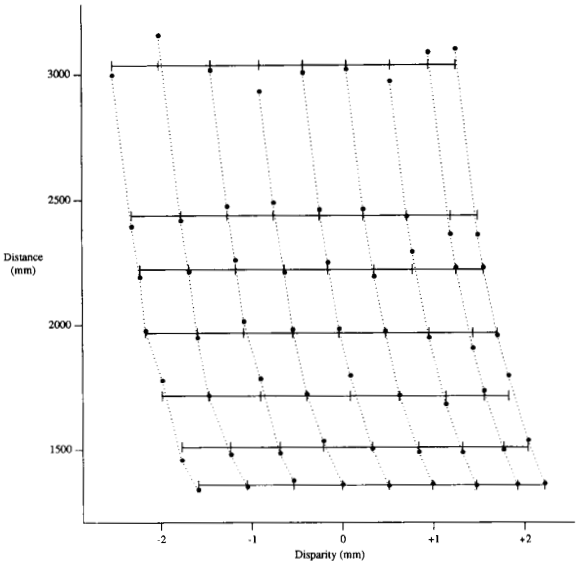


Fig. 6. Distance versus disparity for Scene A. The solid lines connect measured points (drawn as vertical ticks) that lie at the same distance. The dotted lines connect computed points (drawn as large dots) with the same vergence angle. The effect of increasing the convergence angle is to shift the dotted lines to the left along the disparity axis. Note that convergence does not significantly affect the shape of the curves.

TABLE I
Range Errors Using Disparities Computed with Supervision

Scene	Number of Points	μ_E mm	σ_E mm	μ_R percent	σ_R percent
A	63	0.0	41.0	0.0	1.9
B	64	−5.5	44.7	−0.2	2.0

TABLE II
Range Errors Using Disparities Computed with No Supervision

Scene	Number of Points	μ_E mm	σ_E mm	μ_R percent	σ_R percent
A	336	28.0	505.2	2.0	32.5
B	294	−21.1	145.9	−1.6	8.7

TABLE III
Range Errors Using Disparities Computed with Partial Supervision

Scene	Number of Points	μ_E mm	σ_E mm	μ_R percent	σ_R percent
A	319	−1.45	111.6	−0.2	5.0
B	278	−7.0	108.8	−0.6	5.5

VI. Discussion

We have presented a novel method for stereo ranging with verging cameras, and analyzed its performance. We derived the analytical relationship between range, vergence angle, and disparity, and showed how to estimate range given indirect measurements of vergence angle, disparity, and other parameters. We demonstrated the precision of the

range computation with real images to be approximately five percent of the true range. Although this performance is acceptable for many applications, extensions and future work are possible.

We can extend the model relating range to measurable quantities. For instance, we could compensate for radial distortions of the lenses, the optical axes not piercing the image centers, and the tilt of the receptor planes with respect to each other.

Another topic for future work is to eliminate or identify mistaken matches, which, as we have seen, can cause large range errors. This is a difficulty faced by all solutions to the correspondence problem; it is not unique to the matching procedure employed here. One possibility is to incorporate knowledge about the scene, in particular, the minimum and maximum object distances. This constrains the disparities to lie within a certain interval, and therefore bounds the disparity error. A second possibility is to enforce consistency between the range from stereo and the range computed by an independent process, e.g., focus ranging [8]. A third possibility is to tune the matcher to minimize the probability of mistakes, e.g., by increasing the number of neighboring matches required to have similar disparities. Of course, this will tend to decrease the number of matches, particularly in areas of range discontinuities.

Another interesting direction of future research is to model the disparity error distribution (and eventually, range error distribution) by a *p-contaminated* distribution. In such a distribution, a large fraction of the time the disparity error follows a smaller variance distribution of feature localization errors, and a small fraction of the time the disparity error follows a larger variance distribution of mistaken matches. To further describe this distribution, let d^* represent the true disparity of an object point, and d_i represent the disparity computed by the ith execution of the matching algorithm. Suppose the additive error model $d_i = d^* + \delta_i$, where the disparity errors δ_i are identically distributed with a p-contaminated Gaussian probability density $f(\delta)$:

$$f(\delta) = (1-p)N\left(d^*, \sigma_{1-p}^2\right) + pc(\delta),$$

where $p \in [0,1]$, and $c(\delta)$ denotes any density function that is unimodal and symmetric about d^*. As an illustration, we sketch what the p-contaminated disparity distribution might be for our data. Assuming the Gaussian density $N(d^*, \sigma_p^2)$ for c, we estimate roughly that $p = 0.05$, $\sigma_{1-p} \approx 0.07$ mm (2 pixels), and $\sigma_p \approx 1.44$ mm (48 pixels). We emphasize that this is an illustration, not a result.

Appendix A
Image to Camera Coordinate Transformation

In this Appendix, we derive the transformation from the image reference frame in units of pixels to the camera reference frame in metric units. The method applies to solid-state sensors with known geometries.

Define two reference frames, one for the image and one for the camera. The origin of the image reference frame is the upper left-hand corner of the image. The u-axis is aligned with the scan-lines, and the v-axis points to the bottom of the image. Let this frame be left-handed. The origin of the camera reference frame is at the lens center. The x-axis is aligned with the scan-lines, and the y-axis points to the top of the image. Let this frame be right-handed.

We account for scaling, rotation, and translation between the two frames by

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} \alpha & 0 \\ 0 & -\alpha \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} + \begin{bmatrix} -N_x I_x/2 \\ -N_y I_y/2 \\ f \end{bmatrix},$$

where N_x is the number of columns, N_y is the number of scan-lines, I_x is the horizontal interelement spacing (in mm) in the direction of the scan-line, I_y is the vertical interelement spacing, and α is the size in mm of a pixel. Note that this is an improper transformation (det $R = -1$); it changes the handedness from left to right.

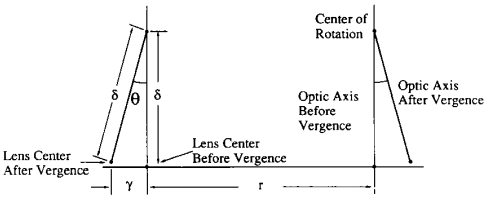


Fig. 7. Vergences changes the baseline distance. This figure illustrates the case where the center of rotation does not coincide with the lens center, but is displaced along the optic axis by an amount δ. When the two optic axes are parallel, the baseline distance is b. When the cameras verge by angle θ, then the baseline distance becomes $b \pm 2\gamma$.

Appendix B
Vergence Changes Baseline Distance

Suppose that the camera stations are not parallel, and that the lens center is displaced from the rotation center by a distance δ along the optic axis. If $\delta > 0$, as in Fig. 7, then the baseline distance b increases as the vergence angle θ increases. Let γ be the distance parallel to the baseline between the rotation center and lens center after vergence. From the triangle in the figure, $\gamma = \delta \sin \theta$. Thus, the baseline distance for verged cameras is $b_v = b + 2\gamma$.

To quantify the change in the baseline caused by vergence, we assign the parameters physically realistic values. The vergence mechanism constrains the vergence angle by $\theta \leq 5°$, and since γ increases with θ, the worst case is $\theta = 5°$. We estimate that $\delta = 1$ cm, implying that $\gamma \approx 0.8$ mm. Under these conditions, the difference 2γ between b_v and b is negligible.

However, since $b_v - b$ increases with δ, the change in baseline distance caused by vergence cannot always be safely neglected.

Acknowledgment

The authors gratefully acknowledge the contributions of the anonymous reviewers, and thank C. Tomasi and S. Nayar for reading drafts.

References

[1] N. Ayache and B. Faverjon, "Efficient registration of stereo images by matching graph descriptions," *Int. J. Comput. Vision,* vol. 1, no. 2, pp. 107–132, 1987.

[2] S. T. Barnard and M. A. Fischler, "Computational stereo," *Comput. Surveys,* vol. 14, no. 4, pp. 553–572, Dec. 1982.

[3] J. B. Burns, A. R. Hanson, and E. M. Riseman, "Extracting straight lines," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. PAMI-8, no. 4, pp. 425–455, July 1986.

[4] U. R. Dhond and J. K. Aggarwal, "Structure from stereo—A review," *IEEE Trans. Syst. Man, Cybern.,* vol. 19, no. 6, pp. 1489–1510, 1989.

[5] O. D. Faugeras and G. Toscani, "The calibration problem for stereo," in *Proc. Computer Vision and Pattern Recognition Conf.,* June 1986, pp. 15–20.

[6] K. Henriksen, "Line based stereo matching," Dep. Comput. Sci., Univ. Copenhagen, Denmark, Tech. Rep. 86-10-10, 1986.

[7] B. K. P. Horn, *Robot Vision.* Cambridge, MA: MIT Press, 1986.

[8] E. Krotkov, "Focusing," *Int. J. Computer Vision,* vol. 1, no. 3, pp. 223–237, Oct. 1987.

[9] E. Krotkov, J. F. Summers, and F. Fuma, "An agile stereo camera system for flexible image acquisition," *IEEE J. Robotics Automation,* vol. 4, no. 1, pp. 108–113, Feb. 1988.

[10] E. Krotkov, *Active Computer Vision by Cooperative Focus and Stereo.* New York: Springer-Verlag, 1989.

[11] I. Sutherland, "Three-dimensional data input by tablet," *Proc. IEEE,* vol. 62, no. 4, pp. 453–461, Apr. 1974.

[12] R. Y. Tsai, "A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE J. Robotics Automation,* vol. RA-3, no. 4, 1987.

[13] Y. Yakimovsky and R. Cunningham, "A system for extracting three-dimensional measurements from a stereo pair of TV cameras," *Comput. Graphics Image Processing,* vol. 7, pp. 195–210, 1978.