COMPUTER VISION AND IMAGE UNDERSTANDING
Vol. 63, No. 3, May, pp. 447–461, 1996
ARTICLE NO. 0034

Analysis of Error in Depth Perception with Vergence and Spatially Varying Sensing*

HOSSEIN SAHABI† AND ANUP BASU‡

Department of Computing Science, University of Alberta, 615 General Services Building, Edmonton, T6G 2H1, Canada

Received December 20, 1993; accepted March 15, 1995

In stereo vision the depth of a 3-D point is estimated based on the position of its projections on the left and right images. The image plane of cameras that produces the images consists of discrete pixels. This discretization of images generates uncertainty in estimation of the depth at each 3-D point. In this paper, we investigate the effect of vergence and spatially varying resolution on the depth estimation error. First, vergence is studied when pairs of stereo images with uniform resolution are used. Then the problem is studied for a stereo system similar to that of humans, in which cameras have high resolution in the center and nonlinearly decreasing resolution toward the periphery. In this paper we are only concerned with error in depth perception, assuming that stereo matching is already done.

1. INTRODUCTION

A significant amount of research has been directed towards the development of systems that are able to perceive the three dimensional (3-D) structure of objects. The 3-D information is essential in many applications such as robotic navigation and medical imaging. Stereo vision is an important method for obtaining depth information from a 3-D scene. In stereo, a pair of cameras provides left and right images of a scene. The depth of each 3-D point is estimated based on the position of its projections in the two images.

In the process of depth recovery from stereo images, three major steps are involved. First, two images are preprocessed; the objective is to identify well-defined features in each image. Second, correspondence is established between features that are projections of the same physical entity in the two images. Many stereo matching algorithms have been proposed [1, 12–14, 17, 20, 21, 23–25, 28]. They are classified, based on the matching primitives, into area-based and feature-based techniques. Area-based methods correspond brightness patterns in two images [14, 24]. These algorithms have several drawbacks which are pointed out in [12, 18]. Feature-based methods match features such as edges [3, 9, 11, 13, 15, 21], and linear edge segments [1, 23]. Finally, the depth of each point is obtained using triangulation.

Cameras have an image plane which consists of a number of discrete picture elements (pixels). In general, these pixels are uniformly arranged in a two-dimensional array according to certain industrial standards. The projection of each 3-D point in the scene is approximated to the nearest pixel—the resulting error is referred to as *discretization error.* In stereo, discretization error generates uncertainty in estimation of depth of each 3-D point. In order to model such uncertainty different approaches have been used, such as discrete tolerance limits [2, 7] and multidimensional probability distribution [22]. Figure 1 illustrates this depth estimation error; for all the points lying in each diamond, the same depth is estimated. This model was first introduced by Matthies and Shafer [22]. As this figure illustrates, the depth estimation error grows with distance. However, the error is not a simple function of distance—the diamonds in Fig. 1 are skewed and oriented.

Optimal vertical and horizontal resolution of stereo cameras for minimizing the depth estimation error has been studied by Basu [4]. In this paper, we deal with an active stereo system. Such a system uses a dynamic pair of cameras which can be tilted (rotated about the horizontal axis) and panned (rotated about vertical axis) independently [10, 26]. For such a device, we investigate the effect of vergence on depth estimation. Initially, we assume that a pair of cameras with uniform resolution is given. We attempt to find the optimal vergence angle for minimum depth uncertainty. Then, since vergence does not have a desirable result in the case of uniform resolution, we use a nonuniform arrangement of pixels similar to the human eye for which vergence leads to minimum depth uncertainty.

* This project was supported in part by grants from the Canadian Natural Sciences and Engineering Research Council (Research and Industrially Oriented Research Grants).

† H. Sahabi is currently with Bell-Northern Research, Ottawa, Canada.
‡ E-mail: anup@cs.ualberta.ca.

1077-3142/96 $18.00

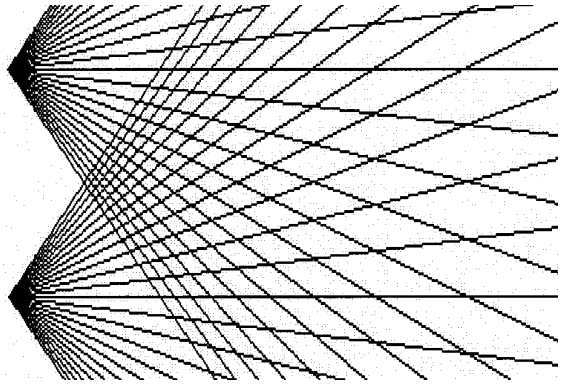

FIG. 1. Depth estimation error in stereo imaging.

Section 2 presents an equation for the depth of a single 3-D point as a function of vergence angle. In Section 3, the effect of vergence on the depth estimation error for a single 3-D point is studied; for a given point with known projections in two images the maximum error is derived. In Section 4, we assume an object of interest is located in the scene; using the results obtained in Section 3, the average depth estimation error is analyzed as a function of the vergence angle. This section also discusses how to minimize the depth estimation error by a combination of changes in the vergence angle and the focal length of the cameras. Section 5 studies the depth estimation error with vergence, using the assumption that the stereo cameras have nonuniform pixel arrangements. Section 6 shows some experimental results.

2. DEPTH OF A 3-D POINT

In this section, we first define a number of terms that are used in this paper. Figure 2 illustrates the configuration

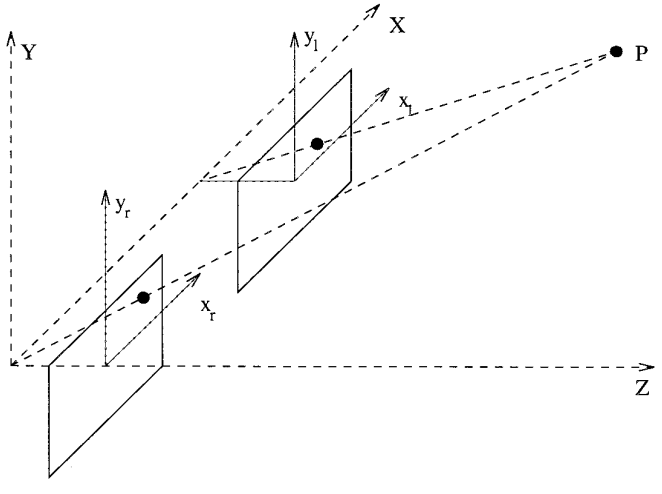

FIG. 2. Stereo imaging system.

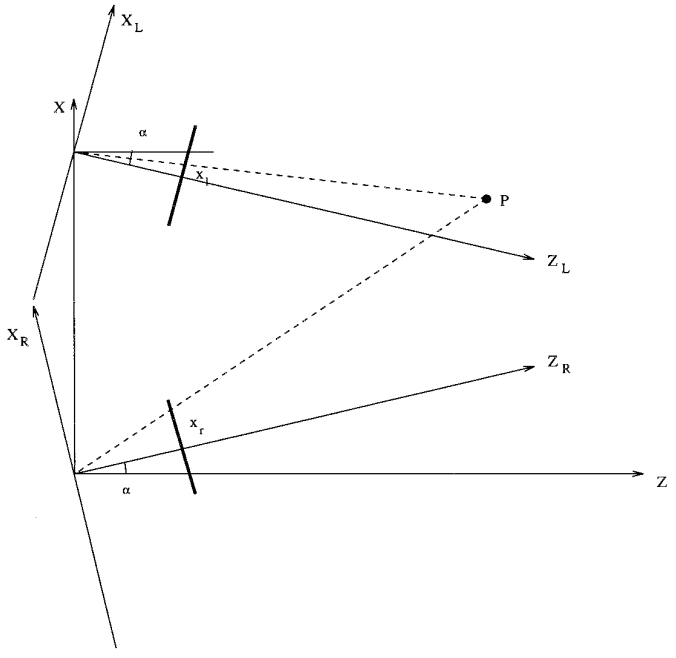

FIG. 3. Stereo imaging system with vergence.

of a stereo imaging system, where two cameras have parallel image planes (no vergence) and are merely separated in the X direction. The focal length of each camera is denoted by f, and the separation distance between the cameras is denoted by dX. It is assumed that the origin of 3-D world coordinates is at the focal point of the right camera. Each camera has a uniform pixel arrangement in both vertical and horizontal directions. The distance between two adjacent pixels along the x direction is de-

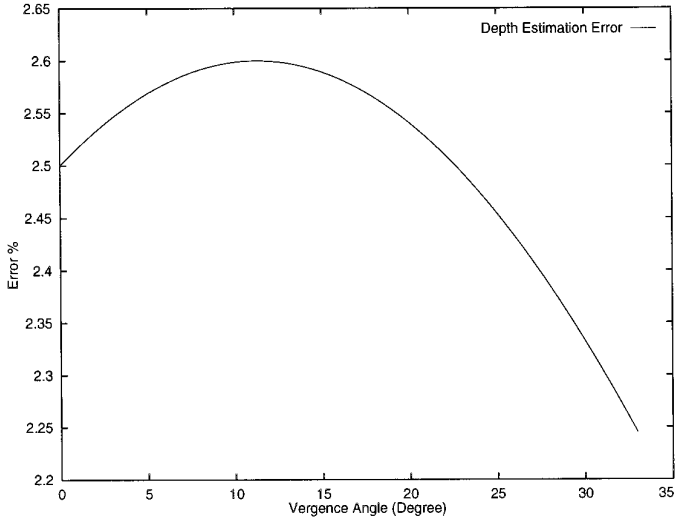

FIG. 4. Maximum relative error in depth estimation versus vergence angle.

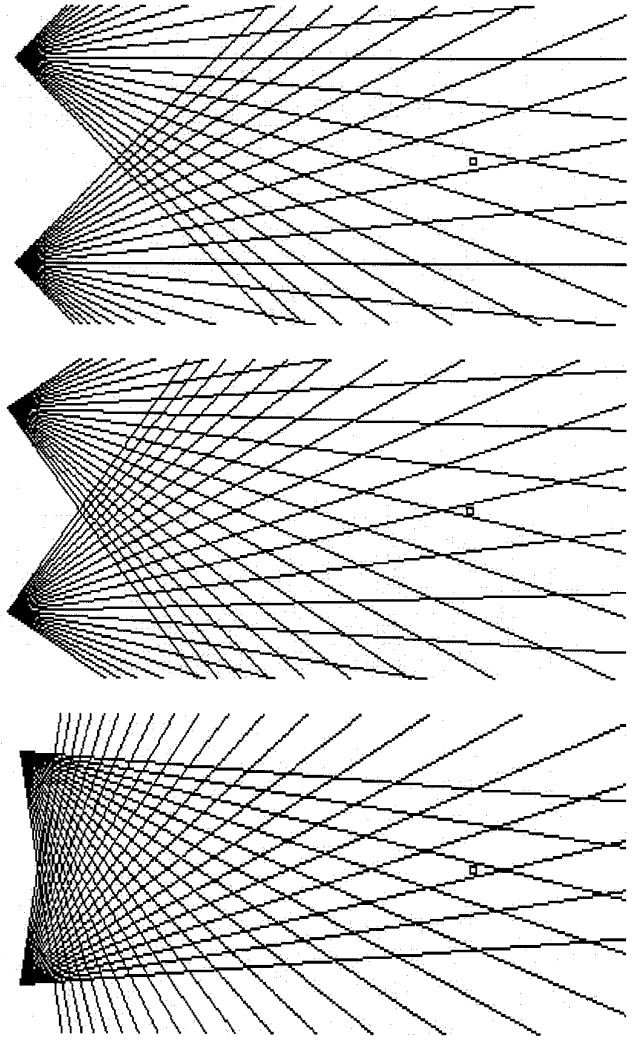

FIG. 5. Variation of depth uncertainty with vergence using uniform resolution images.

noted by e_x, and similarly the distance between two adjacent pixels along the y direction is denoted by e_y. The projection of the 3-D point $P(X, Y, Z)$ in the right and left cameras is represented by (x_r, y_r) and (x_l, y_l), respectively. Because of the discrete placement of pixels, these two projections are approximated to $(\hat{x}_r, \hat{y}_r)$ and $(\hat{x}_l, \hat{y}_l)$. The discretization error in turn leads to an estimate $(\hat{X}, \hat{Y}, \hat{Z})$ of the coordinates of point P.

Figure 3 illustrates a two-dimensional view of a stereo system with vergence angle α. From now on we will not consider the Y coordinate of 3-D points because the Y coordinate does not have any effect on the estimation of depth. The following theorem provides a formula for the calculation of depth (Z coordinate) of the point P in Fig. 3.

THEOREM 1. *The depth of point P is calculated from the equation*

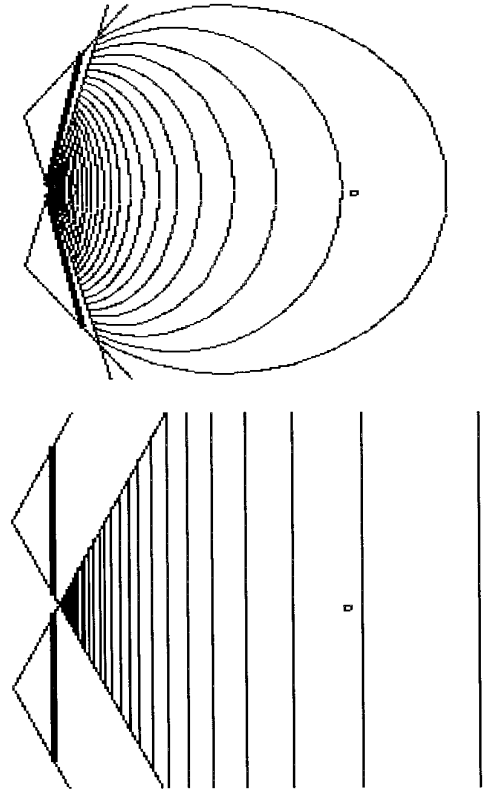

FIG. 6. Isoresolution plots for stereo cameras with uniform discretization.

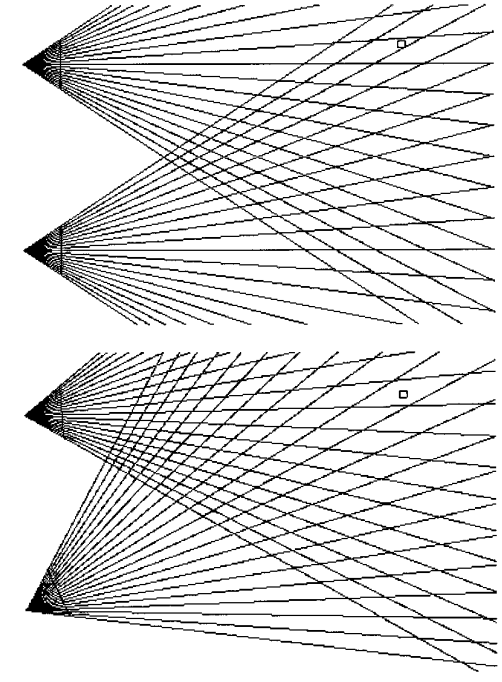

FIG. 7. Variation of depth uncertainty for stereo cameras with uniform discretization and nonsymmetrical vergence angles.

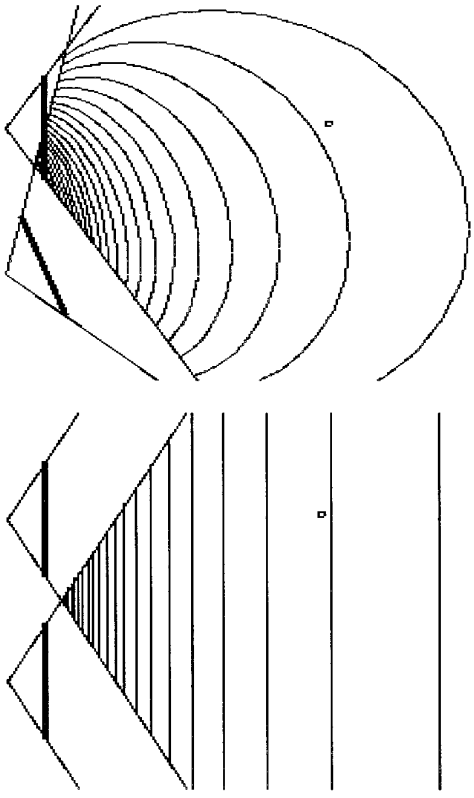

FIG. 8. Isoresolution plots for stereo cameras with uniform discretization and nonsymmetrical vergence angles.

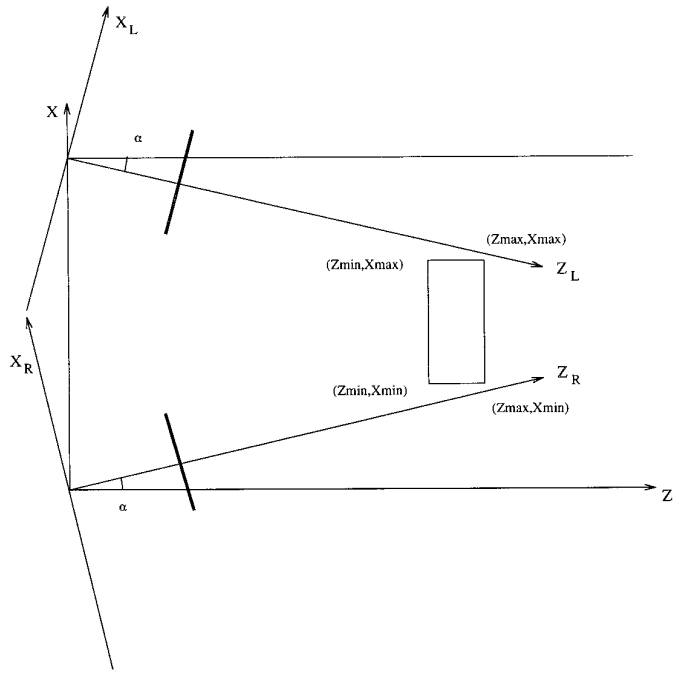

FIG. 9. Stereo system with an object in the scene.

$$Z = \frac{dX(f\cos\alpha + x_l \sin\alpha)(f\cos\alpha - x_r \sin\alpha)}{(f\cos\alpha + x_l \sin\alpha)(f\sin\alpha + x_r \cos\alpha) + (f\sin\alpha - x_l \cos\alpha)(f\cos\alpha - x_r \sin\alpha)}. \quad (1)$$

It should be noted that this theorem does not consider the size of the image plane of the cameras. In reality the size of the image plane is limited and, for certain vergence angles, point P may not lie in the field of view of the cameras. The proof of this theorem is given in the Appendix.

3. OPTIMAL VERGENCE FOR A 3-D POINT

In order to study the effect of vergence on the depth estimation error for a single 3-D point, we consider Theorem 2, which is proved in the Appendix.

THEOREM 2. *The maximum relative error in depth, of point $P(X, Y, Z)$, is given by the equation*

$$E_Z = \left|\frac{\hat{Z} - Z}{Z}\right| \le \frac{e_x \sin\alpha}{2f}\frac{Z\cos\alpha - (X - dX)\sin\alpha}{Z} + \frac{e_x \sin\alpha}{2f}\frac{Z\cos\alpha + X\sin\alpha}{Z} + \frac{e_x \sin\alpha}{2f\,dX}\frac{X(Z\cos\alpha - (X - dX)\sin\alpha)}{Z} + \frac{e_x \cos\alpha}{2f\,dX}(Z\cos\alpha + X\sin\alpha) + \frac{e_x \cos\alpha}{2f\,dX}(Z\cos\alpha - (X - dX)\sin\alpha) - \frac{e_x \sin\alpha}{2f\,dX}\frac{(X - dX)(Z\cos\alpha + X\sin\alpha)}{Z}. \quad (2)$$

According to this theorem, the maximum relative error is directly proportional to e_x. In other words, the higher the resolution of the cameras, the lower the error. Furthermore, the error is inversely proportional to the focal length. Therefore, if the cameras zoom into a point, the error will be reduced.

Now let us consider the behavior of this error with respect to the changes in the vergence angle. Suppose f is 50 mm, the size of image plane is 40×40 mm, dX is 100 mm, and e_x is 0.5 mm. Figure 4 shows the relative depth error versus vergence angle when point P is located at $X = 50$ mm and $Z = 250$ mm. As illustrated in this figure,

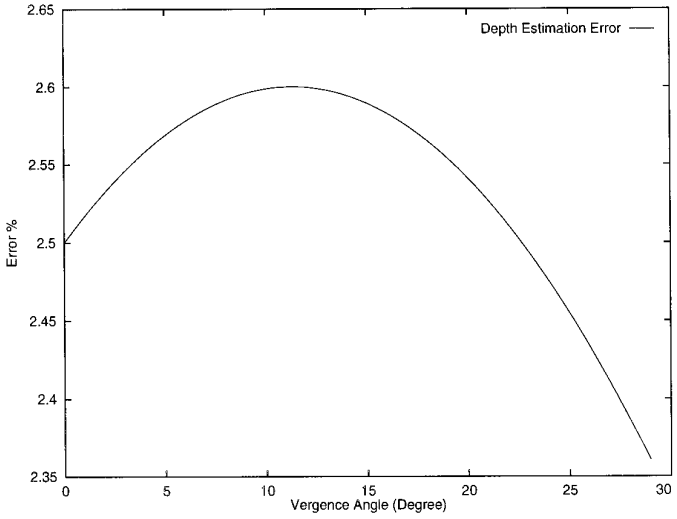

FIG. 10. The average depth estimation error versus vergence angle.

the error increases with vergence; the maximum occurs at almost 11°. Moreover the plot ends at about 33°, which is the angle for which point P goes out of the view of the cameras. In this example the point is symmetrically located with respect to both cameras and will simultaneously be out of the view of the two cameras.

Figure 5 illustrates the variation of depth uncertainty of a point P for three different vergence angles. As illustrated in these figures, the size of the diamond that covers point P changes according to the plot of Fig. 4. Figure 5 demonstrates that the maximum depth estimation error occurs when the projection of the point P is in the center of the image plane of both cameras.

LEMMA 1. *Suppose point P is located symmetrically with respect to two cameras. The vergence angle corresponding to the maximum error for this point is obtained as*

$$\alpha_{\max} = \arctan\left(\frac{dX/2}{Z}\right). \tag{3}$$

Using this lemma with the system in the above example, we have

$$\begin{aligned}\alpha_{\max} &= \arctan\left(\frac{100/2}{250}\right) \\ &= 11.31°.\end{aligned} \tag{4}$$

This result explains the graph in Fig. 4.

Figure 6 illustrates the isoresolution plots for the stereo system with uniform discretization. Each plot represents the points in the scene that have the same maximum error in depth perception. For the zero vergence angle, the isoresolution plots for the points whose depths are not very large compared to dx are almost straight lines. When the vergence angle is increased, the plots become curved. The resolution of the cameras directly affects the distance between the isoresolution plots. The higher the resolution of cameras, the denser the isoresolution plots. The isoresolution plot's density around a fixed point in the scene changes depending on the vergence angle. This verifies the results stated in this section using theoretical studies and the diamond plots.

Figure 7 illustrates the variation of depth uncertainty when stereo cameras have a nonsymmetrical vergence angle. The error in depth perception will be maximum when the projections of a given point in the scene is at the center of the two cameras. This result is also verified by the isoresolution plots in Fig. 8.

4. OPTIMAL VERGENCE FOR AN OBJECT

In this section, we assume that there is an object of interest in the scene. We generalize the results of the previous section, studying the effect of vergence on average depth estimation error, when two constraints are imposed on the location of the object. As depicted in Fig. 9, these constraints are

$$Z\text{min} < Z < Z\text{max}\ X\text{min} < X < X\text{max}.$$

The average depth estimation error is given by the following lemma, which is proved in the Appendix.

LEMMA 2. *The average depth estimation error for the points belonging to an object with the above constraints is*

$$\begin{aligned}\overline{E_z} = \Bigg[&\frac{e_x \sin\alpha\cos\alpha}{f}(X\text{max} - X\text{min})(Z\text{max} - Z\text{min}) \\ &+ \frac{e_x \sin^2\alpha\, dX}{2f}(X\text{max} - X\text{min})\ln\left(\frac{Z\text{max}}{Z\text{min}}\right) \\ &\times \frac{e_x \cos^2\alpha}{2f\,dX}(Z\text{max}^2 - Z\text{min}^2)(X\text{max} - X\text{min}) \\ &+ \frac{e_x \cos\alpha \sin\alpha}{2f}(Z\text{max} - Z\text{min})(X\text{max} - X\text{min}) \\ &- \frac{e_x \sin\alpha\cos\alpha}{2f}(Z\text{max} - Z\text{min})(X\text{max} - X\text{min}) \\ &+ \frac{e_x \sin^2\alpha}{f\,dX}\ln\left(\frac{Z\text{max}}{Z\text{min}}\right) \\ &\times \left(\frac{X\text{max}^3 - X\text{min}^3}{3} - dX\frac{X\text{max}^2 - X\text{min}^2}{2}\right)\Bigg] \\ &/(X\text{max} - X\text{min})(Z\text{max} - Z\text{min}).\end{aligned} \tag{5}$$

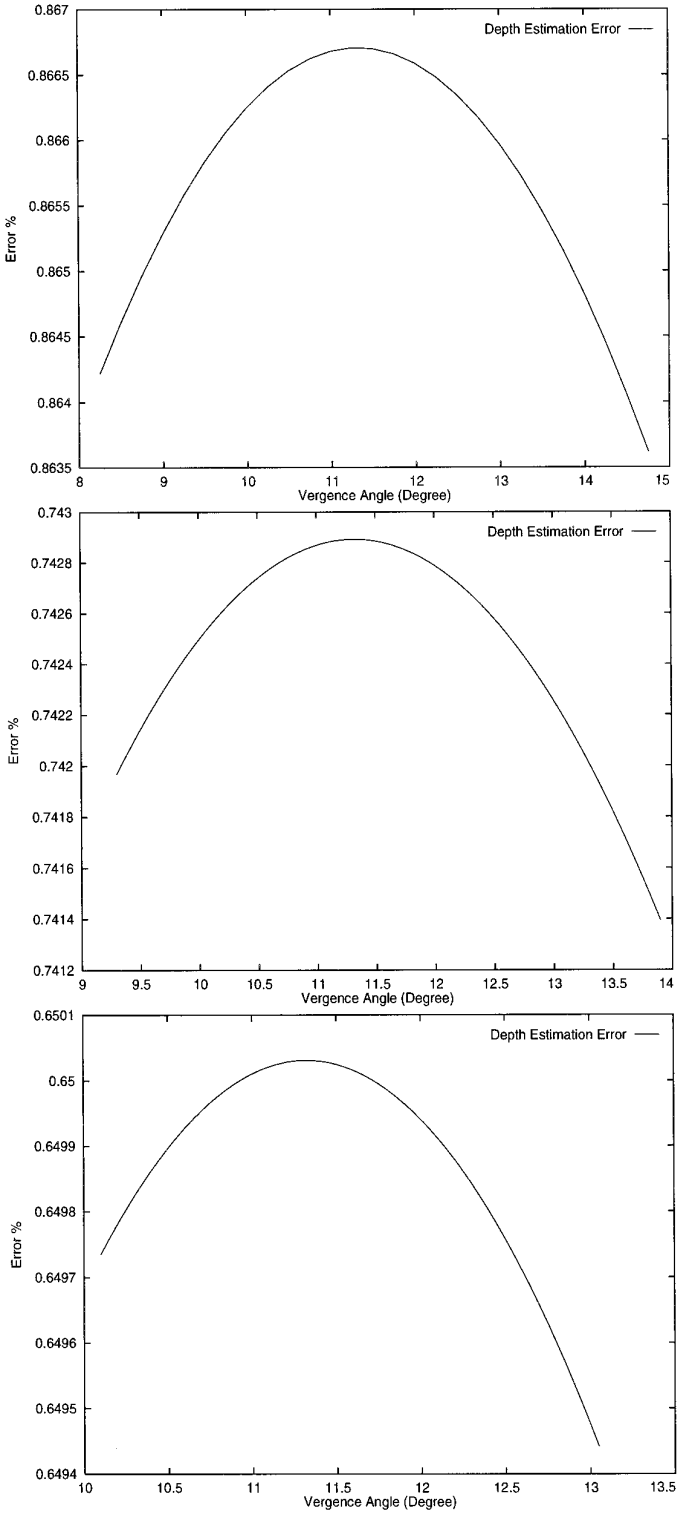

FIG. 11. The average depth estimation error with $f = 150$, 175, and 200.

Suppose an object of interest (in a system with $f = 50$ mm, 40×40 mm imaging area, $e_x = 0.5$ mm, and $dX = 100$ mm) is located in the region of $230 < Z < 270$ and $35 < X < 65$. Figure 10 shows the average error for this object as a function of vergence angle. The maximum error occurs when the center of the object region is projected onto the centers of the cameras. The plot ends when any part of the object is out of view of one of the cameras.

Lemma 2 indicates that the average depth estimation error is inversely proportional to the focal length f. Therefore, we attempt to reduce the depth estimation error by altering both the vergence and focal lengths. Figure 11 illustrates three plots of average depth estimation error for $f = 150$, 175, and 200. By examining these figures, it can be concluded that the average depth estimation error for an object is minimum when the projection of the object covers the whole image plane of the cameras. Such a situation for an object placed symmetrically with respect to two cameras occurs for unique values of f and α.

5. VERGENCE WITH NONUNIFORM DISCRETIZATION

So far we have analyzed the effect of vergence on depth estimation error when the stereo cameras have uniform discretization. As the results of the previous section demonstrate, we can reduce the error by vergence; however, there is no optimal vergence angle which leads to minimum error while maintaining a good coverage of the background scene.

In order to solve this problem, we propose using sensors with nonuniform discretization. In fact we are interested in a stereo system similar to the human visual system. In the human eye, there is a high resolution foveal region in the center, and the resolution decreases toward the periphery. Several approaches for modeling variable resolution images have been proposed [5, 8, 27]. We assume each camera has a nonuniform pixel arrangement similar to the human eye and modeled by the fish-eye transform [6]. The fish-eye transform, which is based on the characteristics of fish-eye lenses, describes a variable resolution mapping of a uniform resolution image to an image with high resolution in the center and nonlinearly decreasing resolution toward the periphery. Based on this transform, any point with coordinates $(\hat{x}, \hat{y})$ in the variable resolution plane is mapped to the point (x, y) in the uniform resolution plane as

$$\rho = \sqrt{\hat{x}^2 + \hat{y}^2} \quad \hat{\theta} = \arctan\left(\frac{\hat{y}}{\hat{x}}\right)$$

$$r = \frac{e^{(\rho/s)-1}}{\lambda} \quad \hat{\theta} = \theta \tag{6}$$

$$x = r\cos\theta \quad y = r\sin\theta.$$

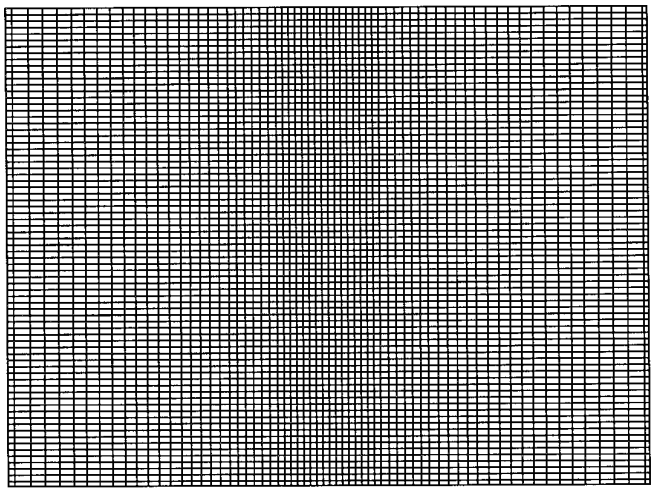

FIG. 12. Exponential pixel distribution in the image plane for $\gamma = 0.03$.

In the above equation s is a simple scaling factor, and λ controls the amount of distortion over the entire range. Using the fish-eye model and ignoring the y coordinate, we derive the following equation for the separation of pixels:

$$e_x(x) = \frac{e^{E/s} - 1}{\lambda} e^{x/s}. \tag{7}$$

In this equation E is a constant which is equal to the separation of pixels in the uniform resolution plane. As is observed from this equation, the separation in this case is exponential with respect to the position of pixels. In fact, Eq. (7) can be rewritten into the form

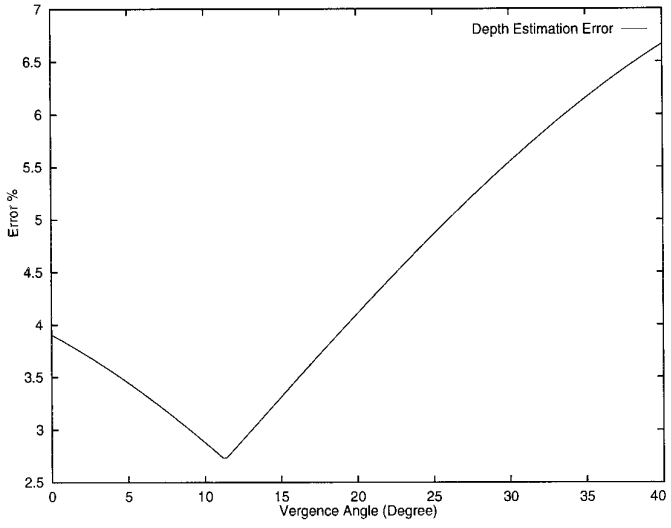

FIG. 13. Depth estimation error when fish-eye images are used.

$$e_x(x) = E_{\min} e^{\gamma x}, \tag{8}$$

where $E_{\min}$ is the smallest pixel separation in the center of the camera, and γ is the factor that determines the rate of increase in the pixel separation with distance. Figure 12 illustrates the pixel distribution for $\gamma = 0.03$. Since e_y does not contribute to the depth estimation error, the pixel separation in the Y direction is uniform.

In order to compute the depth estimation error in a stereo system with nonuniform discretization we use the following lemma.

LEMMA 3. *Suppose we have a single* 3*-D point* $P(X, Y, Z)$ *which has projections* x_r *and* x_l *on the right and left cameras respectively. Each camera has a nonuniform discretization; such a discretization is specified by a function* $e_x(x)$ *which defines the separation of pixels as a function of their location. The maximum relative error in depth estimation of* P *is given by*

$$\begin{aligned} E_Z = \left|\frac{\hat{Z} - Z}{Z}\right| \le{}& \frac{e_x(x_l)\sin\alpha}{2f}\frac{Z\cos\alpha - (X - dX)\sin\alpha}{Z} \\ &+ \frac{e_x(x_r)\sin\alpha}{2f}\frac{Z\cos\alpha + X\sin\alpha}{Z} \\ &+ \frac{e_x(x_l)\sin\alpha}{2f\,dX}\frac{X(Z\cos\alpha - (X - dX)\sin\alpha)}{Z} \\ &+ \frac{e_x(x_r)\cos\alpha}{2f\,dX}(Z\cos\alpha + X\sin\alpha) \\ &+ \frac{e_x(x_l)\cos\alpha}{2f\,dX}(Z\cos\alpha - (X - dX)\sin\alpha) \\ &- \frac{e_x(x_r)\sin\alpha}{2f\,dX}\frac{(X - dX)(Z\cos\alpha + X\sin\alpha)}{Z}. \end{aligned} \tag{9}$$

Suppose in the system described in the preceding sections, we use cameras with a fish-eye pixel arrangement ($E_{\min} = 0.5, \gamma = 0.03$). Usng Lemma 3, the depth estimation error of point P located at $X = 50$ mm and $Z = 250$ mm is illustrated in Fig. 13. The error of P is *minimum* for the same vergence angle at which the error in the uniform resolution system was *maximum*; this is the vergence angle for which the projection of P lies on the center of the image plane of the cameras. Figure 14 illustrates the variation of depth uncertainty of a point P for three different vergence angles and sensors with nonuniform discretization. The plot of Fig. 13 is graphically verified in terms of the size of diamonds in Fig. 14.

Similar to the cameras with uniform resolution, we calculate the average error for an object of interest in the scene. Same constraints on the X and Z dimensions of the object are imposed. In his case, finding an analytic solution for

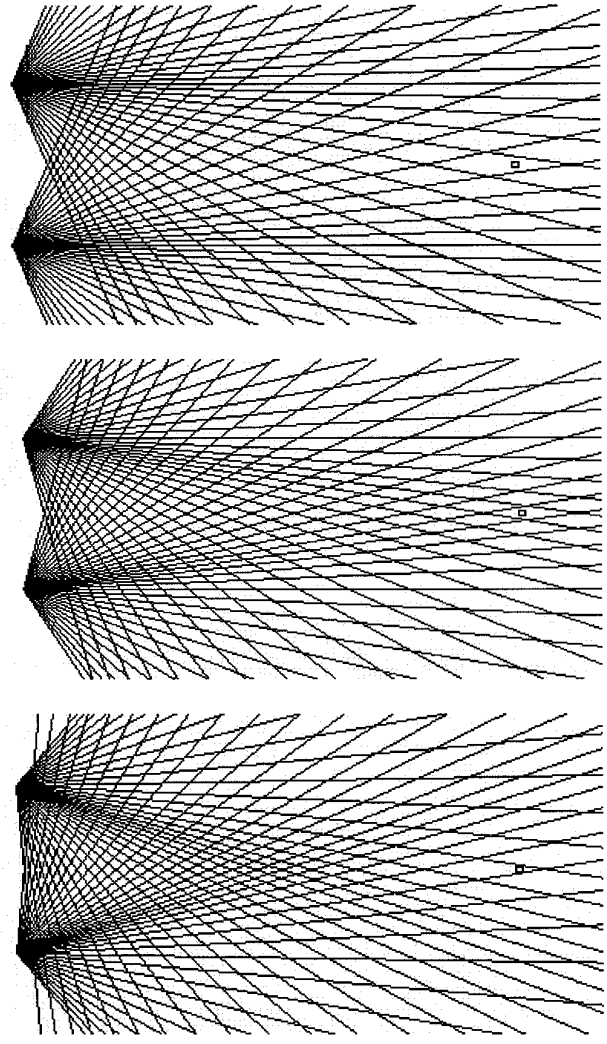

FIG. 14. Variation of depth uncertainty with vergence using fish-eye images.

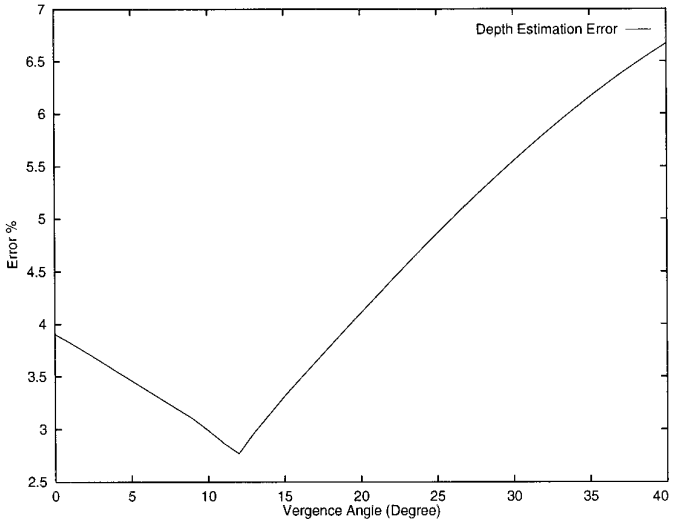

FIG. 15. The average depth estimation error when fish-eye images are used.

the average error is too complicated. Therefore, we solve the following equation using the numerical methods

$$\overline{E_Z} = \frac{\int_{X=X\min}^{X=X\max}\int_{Z=Z\min}^{Z=Z\max} E_Z\, dX\, dZ}{\int_{X=X\min}^{X=X\max}\int_{Z=Z\min}^{Z=Z\max} dX\, dZ}, \qquad (10)$$

where E_Z is obtained from Eq. (9). Suppose the object of interest (in a system with $f = 50$ mm, 40×40 mm imaging area, $e_x = 0.5$ mm, and $dX = 100$ mm) is located in the region of $230 < Z < 270$ and $35 < X < 65$. Figure 15 shows the average error for this object as a function of vergence angle. The minimum error occurs when the center of the object region is projected. The error of depth perception is obviously higher for peripherally visible targets.

Figure 16 illustrates the isoresolution plots for stereo cameras with nonuniform exponential resolution. The density of plots changes with vergence, in accordance with the theoretical results and diamond plots.

Figure 17 illustrates the variation of depth uncertainty when stereo cameras have nonsymmetrical vergence angle. The error in depth perception will be minimum when the projections of the given point in the scene are in the center of the cameras. The result is also verified by the isoresolution plots illustrated in Fig. 18.

One interesting issue that needs investigation is how to select the parameter γ in the nonuniform exponential system such that the error is independent of the vergence angle. By studying the diamond plots from the uniform resolution system, we realize that the behavior of depth perception error with vergence is directly related to the focal angles[1] of the pixels. For the uniform resolution cameras, the focal angle of the pixels in the center of the image plane are larger than the focal angle of pixels in the periphery. This is the reason why the depth perception error is larger when the vergence angle is selected such that the projection of the 3-D point is in the center of the camera. For the nonuniform exponential resolution, there is no value for parameter γ that results in equal focal angles for all the pixels, and, therefore, the error plot versus vergence is never flat. However, the parameter γ should be chosen such that the focal angle of pixels increases monotonically, going from the center of an image plane

[1] If the left and right boundaries of a pixel are connected by lines to the focal point of the camera, the two resulting lines create an angle which is named the focal angle of the pixel. This concept is clearly illustrated in the diamond plots in this paper.

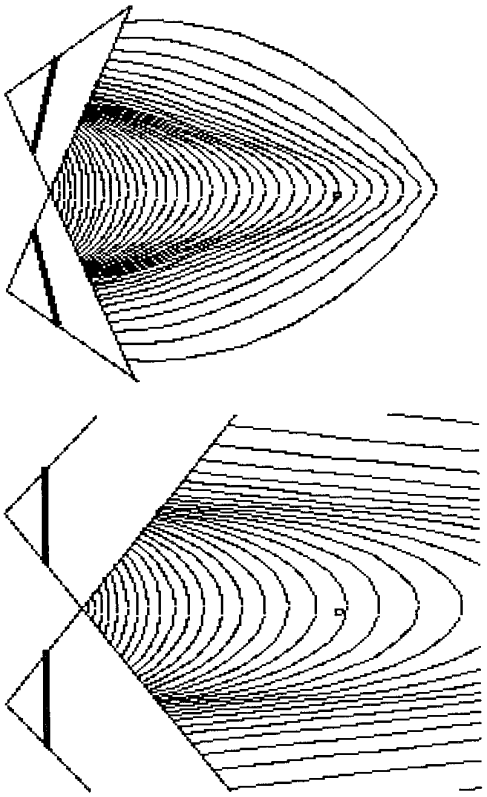

FIG. 16. Isoresolution plots for stereo cameras with nonuniform discretization.

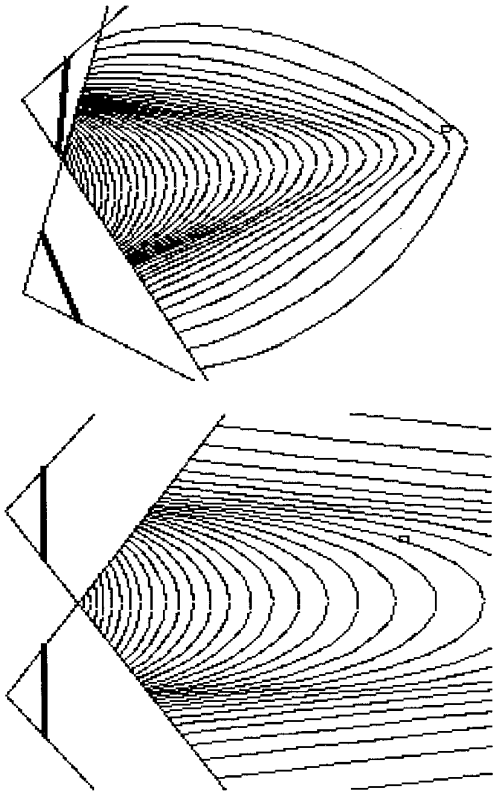

FIG. 18. Isoresolution plots for stereo cameras with uniform discretization and nonsymmetrical vergence angles.

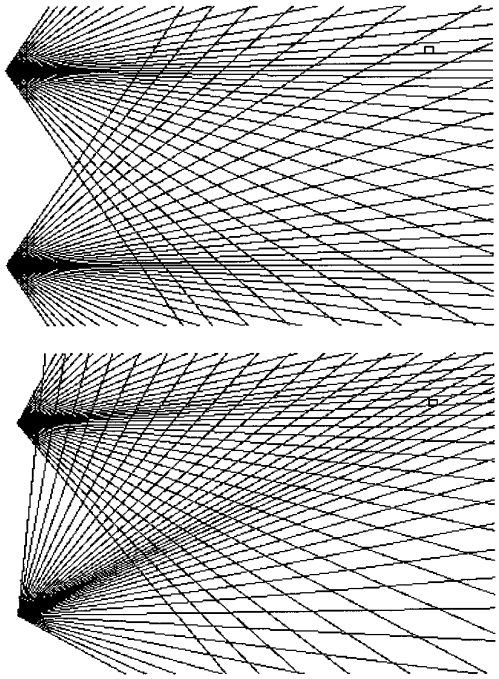

FIG. 17. Variation of depth uncertainty for stereo cameras with nonuniform discretization, and nonsymmetrical vergence angles.

toward the periphery. For small values of λ, although the pixel separation increases monotonically from the center toward the periphery, the focal angles may initially decrease and then increase—resulting in depth perception error changing in a nonmonotonic fashion. The following lemma provides the minimum value of γ that results in a monotonic increase of focal angles from the center toward

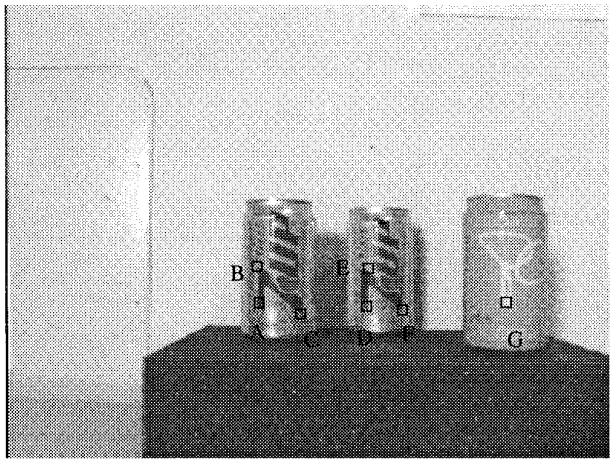

FIG. 19. Points at which depth was estimated.

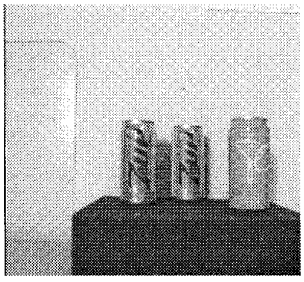

FIG. 20. Stereo images with no vergence.

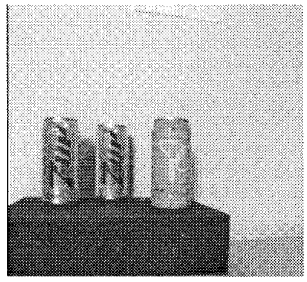

FIG. 22. Stereo images with 16° vergence.

the periphery. The important point to note here is that the value of λ is derived from the geometry of the camera and not from the characteristics of the scene.

LEMMA 4. *The minimum value of γ for monotonic increase in focal angles of pixels from the center of image plane toward the periphery is*

$$\gamma_{\min} = \ln\left(\frac{2f^2}{f^2 - E_{\min}^2} - 1\right). \tag{11}$$

6. EXPERIMENTAL RESULTS

This section presents the results of a number of experiments which were conducted to verify the theoretical results achieved in the previous sections. The error of depth of seven 3-D points, marked in Fig. 19, was evaluated using several pairs of stereo images. Stereo images were acquired using a single CCD camera ($f = 8.37018$ mm) at two different positions ($dX = 112$ mm); the camera was moved horizontally from the first position to the second. The vergence angle was measured precisely from the amount of displacement of a particular point in the image. Points from the left and right images were correlated manually with high accuracy.

In the first set of experiments, three pairs of uniform resolution stereo images (128×120) acquired with three different vergence angles were used, as illustrated in Figs. 20 to 22. Table 1 shows the maximum percentage of relative error of the estimated depth at each point, for different vergence angles. From the results of this table, we observe that the error of depth estimation of different points varies in a different manner. The maximum error of points D, E, and F occurs at 6.36° vergence, the reason being that the projections of these points are closer to the center of the images. For the other points, which are situated asymmetrically with respect to two cameras, the error varies with vergence; however the maximum does not occur at the same vergence as points D, E, and F.

In the second set of experiments, the error of depth was evaluated using fish-eye images, as shown in Figs. 23 and 24. Fish-eye images were acquired by transforming high resolution (512×480) images. The resulting fish-eye images were inversely transformed to the uniform resolution plane; however, the resolution in the periphery of the final images was reduced, while the resolution in the center was preserved. In contrast to the uniform resolution case, the error of points D, E, and F is minimum at 6.36° vergence (Table 2). For the other points, the minimum error does not necessarily occur at the same vergence angle.

In the above experiments, we illustrated how the maximum error in the depth of the selected points in the scene varies with the vergence angle. The results of experiments for individual points are not meaningful, since the discretization error resulting from projection of a single point can

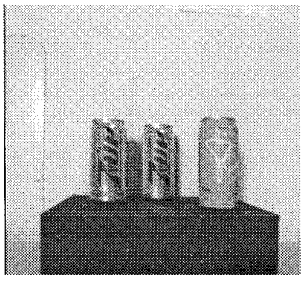

FIG. 21. Stereo images with 6° vergence.

TABLE 1

Maximum Percentage of Error in Depth of the Selected Points in the Scene Using Stereo Images with Uniform 128×120 Resolution and Different Vergence Angles

Point	$\alpha = 0$	$\alpha = 6.36$	$\alpha = 16.32$
A	3.70	3.75	3.70
B	3.70	3.74	3.67
C	3.70	3.70	3.74
D	4.17	4.21	4.04
E	4.17	4.21	4.04
F	4.17	4.19	4.01
G	3.13	3.24	3.21

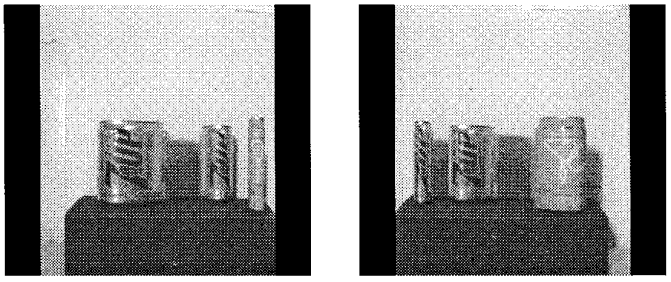

FIG. 23. Fish-eye stereo images with no vergence.

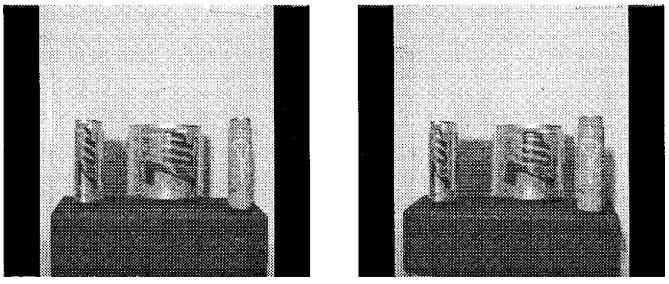

FIG. 24. Fish-eye stereo images with 6° vergence.

vary significantly. We therefore consider the average error for a number of discrete points lying on an artificial object. We choose an artificial scene for convenience. For a real scene, points will have to be selected manually, which can be a tedious process.

In the first experiment on the artificial scene, we assume the stereo cameras have uniform discretization. In this experiment we assume that $e_x = 0.5$, the focal length (f) of each camera is 50, and the distance (dx) is 100 mm. The average error (EZ) in the depth is calculated for points lying on a plane with depth $Z = 350$ and on a grid size of 1 mm within the boundaries $25 < X < 75$ and $50 < Y < 100$. Figure 25 illustrates the changes in average value of EZ with the vergence angle.

In the second experiment on the artificial scene, we assume that the stereo cameras have nonuniform exponential discretization. In this experiment we assume that $E_{\min} = 0.5$, $\gamma = 0.05$, the focal length (f) of each camera is 50, and the distance (dx) is 100 mm. The average error (EZ) in the depth is calculated for points lying on a plane with depth $Z = 350$ and on a grid size of 1 mm within the boundaries $25 < X < 75$ and $50 < Y < 100$. Figure 26 illustrates the changes in average value of EZ with the vergence angle.

7. CONCLUSION AND FUTURE WORK

In this paper, we have addressed the problem of determining the optimal vergence for minimum depth estimation error. In the uniform resolution plane, there is no vergence angle for which the error of a particular point is minimum while maintaining an appropriate view of the background. When an object of interest is located in the scene, the error is minimized with a combination of changes in vergence and focal length. We also considered vergence in a stereo system similar to the human eyes. In this case, cameras have higher resolution in the center and reduced resolution in the periphery. Vergence in this variable resolution stereo system has desirable effects. Error in depth estimate is minimized when both cameras "look at" an object of interest.

In this paper, we were only concerned with the analysis of error in depth perception assuming that stereo correspondence is given. The results are particularly important for designing optimal stereo displays—in this case the stereo correspondence is done by the human brain. In future research, we will consider the problem of stereo correspondence for stereo images with spatially varying resolution. We will also study a more general active stereo system in which each camera has a combination of pan,

TABLE 2
Maximum Percentage of Error in Depth of the Selected Points in the Scene Using Stereo Fish-Eye Images ($\gamma = 0.0135$) and Different Vergence Angles

Point	$\alpha = 0$	$\alpha = 6.36$	$\alpha = 16.32$
A	1.37	1.28	1.34
B	1.43	1.38	1.29
C	1.12	0.94	1.18
D	0.96	0.06	1.32
E	0.93	0.05	1.37
F	1.00	0.08	1.35
G	1.31	1.26	1.22

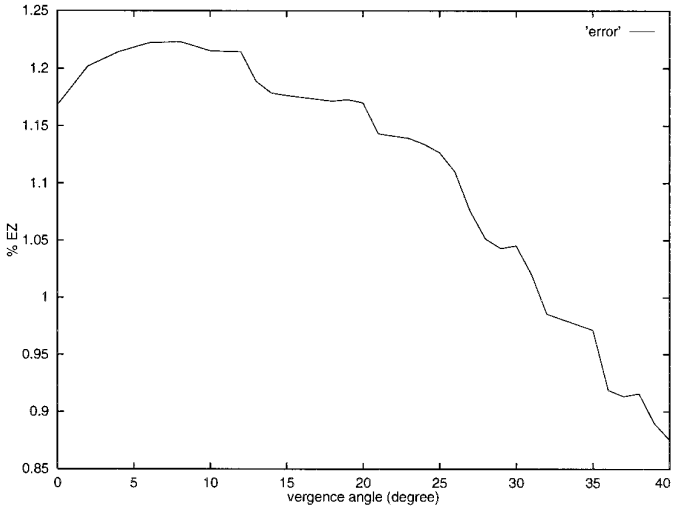

FIG. 25. The average error in depth estimation versus vergence angle for the artificial scene, when stereo cameras have uniform discretization.

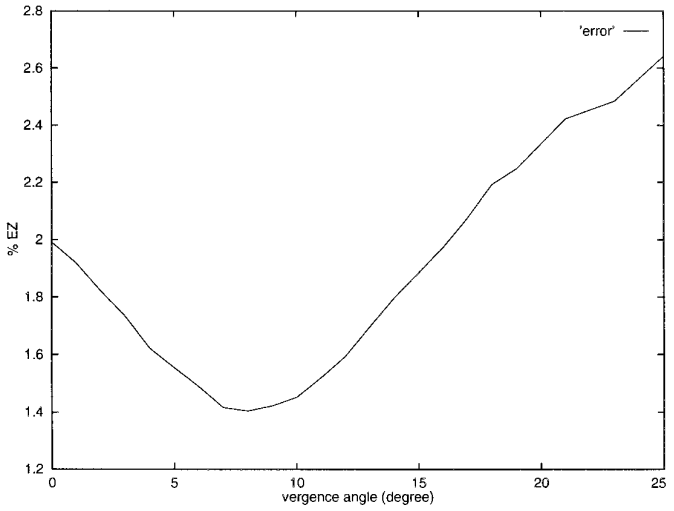

FIG. 26. The average error in depth estimation versus vergence angle for the artificial scene, when stereo cameras have nonuniform exponential discretization.

tilt, and torsional movements [16] and are not necessarily symmetrical like the human visual system.

APPENDIX

THEOREM 1. *Depth of point P is calculated from the equation*

$$Z = \frac{dX(f\cos\alpha + x_l\sin\alpha)(f\cos\alpha - x_r\sin\alpha)}{(f\cos\alpha + x_l\sin\alpha)(f\sin\alpha + x_r\cos\alpha) + (f\sin\alpha - x_l\cos\alpha)(f\cos\alpha - x_r\sin\alpha)}.$$

Proof. In order to prove this theorem, a new coordinate system for each camera is defined, as illustrated in Fig. 3. The origin of the new coordinate system is located in the focal point of the camera and its Z axis is perpendicular to the image plane. Let (X_R, Y_R, Z_R) and (X_L, Y_L, Z_L) be the coordinates of P in the new coordinate systems, which correspond to the right and left cameras respectively. The following relations between the projection of P and its coordinates in each of these systems exist:

$$x_r = \frac{fX_R}{Z_R} \quad x_l = \frac{fX_L}{Z_L}. \tag{12}$$

X_R, Z_R, X_L, and Z_L can be converted to X and Z (the main 3-D coordinate system) using the equations

$$\begin{aligned} Z_R &= Z\cos\alpha + X\sin\alpha \\ X_R &= X\cos\alpha - Z\sin\alpha \end{aligned} \tag{13}$$

$$\begin{aligned} Z_L &= Z\cos\alpha - (X - dX)\sin\alpha \\ X_L &= (X - dX)\cos\alpha + Z\sin\alpha. \end{aligned} \tag{14}$$

Using these relationships, Eq. (12) is rewritten as

$$\begin{aligned} x_r &= \frac{f(X\cos\alpha - Z\sin\alpha)}{(Z\cos\alpha + X\sin\alpha)} \\ x_l &= \frac{f((X - dX)\cos\alpha + Z\sin\alpha))}{(Z\cos\alpha - (X - dX)\sin\alpha)}. \end{aligned} \tag{15}$$

By eliminating X from these equations, we have

$$Z = \frac{dX(f\cos\alpha + x_l\sin\alpha)(f\cos\alpha - x_r\sin\alpha)}{(f\cos\alpha + x_l\sin\alpha)(f\sin\alpha + x_r\cos\alpha) + (f\sin\alpha - x_l\cos\alpha)(f\cos\alpha - x_r\sin\alpha)}.$$

THEOREM 2. *The maximum relative error in depth of point $P(X, Y, Z)$ is obtained by the equation*

$$\begin{aligned} E_Z = \left|\frac{\hat{Z} - Z}{Z}\right| \le{}& \frac{e_x\sin\alpha}{2f}\frac{Z\cos\alpha - (X - dX)\sin\alpha}{Z} \\ &+ \frac{e_x\sin\alpha}{2f}\frac{Z\cos\alpha + X\sin\alpha}{Z} \\ &+ \frac{e_x\sin\alpha}{2fdX}\frac{X(Z\cos\alpha - (X - dX)\sin\alpha)}{Z} \\ &+ \frac{e_x\cos\alpha}{2fdX}(Z\cos\alpha + X\sin\alpha) \\ &+ \frac{e_x\cos\alpha}{2fdX}(Z\cos\alpha - (X - dX)\sin\alpha) \\ &- \frac{e_x\sin\alpha}{2fdX}\frac{(X - dX)(Z\cos\alpha + X\sin\alpha)}{Z}. \end{aligned}$$

Proof. The projection of point P in each camera has at most $e_x/2$ discretization error. We have

$$\hat{x}_r = x_r \pm e_x/2 \quad \hat{x}_l = x_l \pm e_x/2. \tag{16}$$

Using Theorem 1, the depth of point P is estimated as

$$\hat{Z} = \frac{dX(f\cos\alpha + \hat{x}_l\sin\alpha)(f\cos\alpha - \hat{x}_r\sin\alpha)}{(f\cos\alpha + \hat{x}_l\sin\alpha)(f\sin\alpha + \hat{x}_r\cos\alpha) + (f\sin\alpha - \hat{x}_l\cos\alpha)(f\cos\alpha - \hat{x}_r\sin\alpha)}. \tag{17}$$

By substituting for $\hat{x}_r$ and $\hat{x}_l$ in the above equation, we have

$$\begin{aligned}\hat{Z} = dX\Bigg(f\cos\alpha + x_l\sin\alpha \pm \frac{e_x\sin\alpha}{2}\Bigg)&\\ \Bigg(f\cos\alpha - x_r\sin\alpha \pm \frac{e_x\sin\alpha}{2}\Bigg)&\\ \Bigg/\Bigg[\Bigg(f\cos\alpha + x_l\sin\alpha \pm \frac{e_x\sin\alpha}{2}\Bigg)&\\ \Bigg(f\sin\alpha + x_r\cos\alpha \pm \frac{e_x\cos\alpha}{2}\Bigg)&\\ + \Bigg(f\sin\alpha - x_l\cos\alpha \mp \frac{e_x\cos\alpha}{2}\Bigg)&\\ \Bigg(f\cos\alpha - x_r\sin\alpha \mp \frac{e_x\sin\alpha}{2}\Bigg)\Bigg]&\\ = \mathrm{Num}(Z)\Bigg(1 \pm \frac{e_x\sin\alpha}{2(f\cos\alpha + x_l\sin\alpha)}\Bigg)&\\ \Bigg(1 \pm \frac{e_x\sin\alpha}{2(f\cos\alpha - x_r\sin\alpha)}\Bigg)\Bigg/(\mathrm{Den}(Z) \pm G(Z)),&\end{aligned} \tag{18}$$

where $\mathrm{Num}(Z)$ and $\mathrm{Den}(Z)$ represent numerator and denominator of the equation of Theorem 1, and $G(Z)$ is

$$\begin{aligned}G(Z) = &\frac{e_x\sin\alpha}{2}(f\sin\alpha + x_r\cos\alpha)\\ &+ \frac{e_x\cos\alpha}{2}(f\cos\alpha + x_l\sin\alpha)\\ &+ \frac{e_x\cos\alpha}{2}(f\cos\alpha - x_r\sin\alpha)\\ &+ \frac{e_x\sin\alpha}{2}(f\sin\alpha - x_l\cos\alpha).\end{aligned} \tag{19}$$

By replacing $Z/\mathrm{Num}(Z)$ for $\mathrm{Den}(Z)$, Eq. (18) will be simplified to

$$\begin{aligned}\hat{Z} = Z\Bigg(1 \pm \frac{e_x\sin\alpha}{2(f\cos\alpha + x_l\sin\alpha)}\Bigg)\Bigg(1 \pm \frac{e_x\sin\alpha}{2(f\cos\alpha - x_r\sin\alpha)}\Bigg)&\\ \Bigg/\Bigg(1 \pm Z\frac{G(Z)}{dX(f\cos\alpha + x_l\sin\alpha)(f\cos\alpha - x_r\sin\alpha)}\Bigg).&\end{aligned} \tag{20}$$

For practical situations where Z is not very large,

$$Z\frac{G(Z)}{dX(f\cos(\alpha) + x_l\sin(\alpha))(f\cos(\alpha) - x_r\sin(\alpha))}$$

is small and the above equation is rewritten into the form

$$\begin{aligned}\hat{Z} \cong Z\Bigg(1 \pm \frac{e_x\sin\alpha}{2(f\cos\alpha + x_l\sin\alpha)}\Bigg)&\\ \times\Bigg(1 \pm \frac{e_x\sin\alpha}{2(f\cos\alpha - x_r\sin\alpha)}\Bigg)&\\ \times\Bigg(1 \mp Z\frac{G(Z)}{dX(f\cos\alpha + x_l\sin\alpha)(f\cos\alpha - x_r\sin\alpha)}\Bigg).&\end{aligned} \tag{21}$$

By the expansion of the right-hand side of the above equation, some of the terms that are the result of multiplication of two or three fractional terms in the parentheses are negligible. Thus,

$$\begin{aligned}\hat{Z} \cong Z\Bigg(1 \pm \frac{e_x\sin\alpha}{2(f\cos\alpha + x_l\sin\alpha)} \pm \frac{e_x\sin\alpha}{2(f\cos\alpha - x_r\sin\alpha)}&\\ \mp Z\frac{G(Z)}{dX(f\cos\alpha + x_l\sin\alpha)(f\cos\alpha - x_r\sin\alpha)}\Bigg).&\end{aligned} \tag{22}$$

By substituting $G(Z)$ from Eq. (19) the worst case relative error of Z is given by

$$\begin{aligned}\left|\frac{\hat{Z} - Z}{Z}\right| \le \Bigg\{&\frac{e_x\sin\alpha}{2(f\cos\alpha + x_l\sin\alpha)} + \frac{e_x\sin\alpha}{2(f\cos\alpha - x_r\sin\alpha)}\\ &\times\frac{e_x\sin\alpha Z(f\sin\alpha + x_r\cos\alpha)}{2dX(f\cos\alpha + x_l\sin\alpha)(f\cos\alpha - x_r\sin\alpha)}\\ &+ \frac{e_x\cos\alpha Z}{2dX(f\cos\alpha - x_r\sin\alpha)}\\ &+ \frac{e_x\cos\alpha Z}{2dX(f\cos\alpha + x_l\sin\alpha)}\\ &+ \frac{e_x\sin\alpha Z(f\sin\alpha - x_l\cos\alpha)}{2dX(f\cos\alpha + x_l\sin\alpha)(f\cos\alpha - x_r\sin\alpha)}\Bigg\}.\end{aligned} \tag{23}$$

Using Eq. (15), the above equation is rewritten as

$$\begin{aligned}E_Z = \left|\frac{\hat{Z} - Z}{Z}\right| \le &\frac{e_x\sin\alpha}{2f}\frac{Z\cos\alpha - (X - dX)\sin\alpha}{Z}\\ &+ \frac{e_x\sin\alpha}{2f}\frac{Z\cos\alpha + X\sin\alpha}{Z}\\ &+ \frac{e_x\sin\alpha}{2fdX}\frac{X(Z\cos\alpha - (X - dX)\sin\alpha)}{Z}\\ &+ \frac{e_x\cos\alpha}{2fdX}(Z\cos\alpha + X\sin\alpha)\end{aligned}$$

$$+\frac{e_x\cos\alpha}{2fdX}(Z\cos\alpha-(X-dX)\sin\alpha)$$
$$-\frac{e_x\sin\alpha}{2fdX}\frac{(X-dX)(Z\cos\alpha+X\sin\alpha)}{Z}.$$

LEMMA 2. *The average depth estimation error for the points belonging to an object with the above constraints is computed from*

$$\overline{E_Z}=\Bigg[\frac{e_x\sin\alpha\cos\alpha}{f}(X\max-X\min)(Z\max-Z\min)$$
$$+\frac{e_x\,dX\sin^2\alpha}{2f}(X\max-X\min)\ln\left(\frac{Z\max}{Z\min}\right)$$
$$\times\frac{e_x\cos^2\alpha}{2fdX}(Z\max^2-Z\min^2)(X\max-X\min)$$
$$+\frac{e_x\cos\alpha\sin\alpha}{2f}(Z\max-Z\min)(X\max-X\min)$$
$$-\frac{e_x\sin\alpha\cos\alpha}{2f}(Z\max-Z\min)(X\max-X\min)$$
$$\pm\frac{e_x\sin^2\alpha}{fdX}\ln\left(\frac{Z\max}{Z\min}\right)$$
$$\left(\frac{X\max^2-X\min^3}{3}-dX\frac{X\max^2-X\min^2}{2}\right)\Bigg]$$
$$/(X\max-X\min)(Z\max-Z\min).$$

Proof. The average depth estimation error is calculated as

$$\overline{E_Z}=\frac{\int_{X=X\min}^{X=X\max}\int_{Z=Z\min}^{Z=Z\max}E_Z\,dX\,dZ}{\int_{X=X\min}^{X=X\max}\int_{Z=Z\min}^{Z=Z\max}dX\,dZ}. \tag{24}$$

Using Theorem 2, we have

$$I(Z)=\int_{X=X\min}^{X=X\max}\int_{Z=Z\min}^{Z=Z\max}E_Z\,dX\,dZ=\int_{X=X\min}^{X=X\max}\int_{Z=Z\min}^{Z=Z\max}$$
$$\times\left\{\frac{e_x\sin\alpha}{2f}\frac{Z\cos\alpha-(X-dX)\sin\alpha}{Z}\right.$$
$$+\frac{e_x\sin\alpha}{2f}\frac{Z\cos\alpha+X\sin\alpha}{Z}$$
$$+\frac{e_x\sin\alpha}{2fdX}\frac{X(Z\cos\alpha-(X-dX)\sin\alpha)}{Z} \tag{25}$$
$$+\frac{e_x\cos\alpha}{2fdX}(Z\cos\alpha+X\sin\alpha)$$
$$+\frac{e_x\cos\alpha}{2fdX}(Z\cos\alpha-(X-dX)\sin\alpha)$$
$$\left.-\frac{e_x\sin\alpha}{2fdX}\frac{(X-dX)(Z\cos\alpha+X\sin\alpha)}{Z}\right\}dX\,dZ$$

$$I(Z)=\frac{e_x\sin\alpha\cos\alpha}{f}(X\max-X\min)(Z\max-Z\min)$$
$$+\frac{e_x\,dX\sin^2\alpha}{2f}(X\max-X\min)\ln\left(\frac{Z\max}{Z\min}\right)$$
$$\times\frac{e_x\cos^2\alpha}{2fdX}(Z\max^2-Z\min^2)(X\max-X\min)$$
$$+\frac{e_x\cos\alpha\sin\alpha}{2f}(Z\max-Z\min)(X\max-X\min)$$
$$+\frac{e_x\sin\alpha\cos\alpha}{2f}(Z\max-Z\min)(X\max-X\min)$$
$$+\frac{e_x\sin^2\alpha}{fdX}\ln\left(\frac{Z\max}{Z\min}\right)$$
$$\times\left(\frac{X\max^3-X\min^3}{3}-dX\frac{X\max^2-X\min^2}{2}\right). \tag{26}$$

Consequently, we have

$$\overline{E_Z}=\Bigg[\frac{e_x\sin\alpha\cos\alpha}{f}(X\max-X\min)(Z\max-Z\min)$$
$$+\frac{e_x\sin^2\alpha\,dX}{2f}(X\max-X\min)\ln\left(\frac{Z\max}{Z\min}\right)$$
$$\times\frac{e_x\cos^2\alpha}{2fdX}(Z\max^2-Z\min^2)(X\max-X\min)$$
$$+\frac{e_x\cos\alpha\sin\alpha}{2f}(Z\max-Z\min)(X\max-X\min)$$
$$-\frac{e_x\sin\alpha\cos\alpha}{2f}(Z\max-Z\min)(X\max-X\min)$$
$$+\frac{e_x\sin^2\alpha}{fdX}\ln\left(\frac{Z\max}{Z\min}\right)$$
$$\times\left(\frac{X\max^3-X\min^3}{3}-dX\frac{X\max^2-X\min^2}{2}\right)\Bigg]$$
$$/(X\max-X\min)(Z\max-Z\min).$$

REFERENCES

1. N. Ayache and B. Faverjon, Efficient registration of stereo images by matching graph descriptions of edge segments, *Int. J. Comput. Vision,* 1987, 107–131.

2. H. S. Baird, *Model-Based Image Matching Using Location,* MIT Press, Cambridge, MA, 1985.
3. D. A. Ballard and C. M. Brown, *Computer Vision,* Prentice–Hall, Englewood Cliffs, NJ, 1982.
4. A. Basu, Optimal discretization for stereo reconstruction, *Pattern Recognit. Lett.* **13**(11), 1992, 813–820.
5. A. Basu and X. Li, A framework for variable resolution vision, in *Advances in Computing Information—ICCI 91, Ottawa, Canada, May 1991,* pp. 721–732.
6. A. Basu and S. Licardie, Modeling fish-eye lenses, in *IEEE IROS Conference, Yokohama, Japan, July 1993.*
7. R. A. Brooks, Symbolic reasoning among 3-D models and 2-D images, *Artif. Intell.* **17,** 1981, 285–348.
8. A. Califano, R. Kjeldsen, and R. M. Bolle, Data and model driven foveation, in *Proceedings of IEEE International Conference on Pattern Recognition,* June 1990, pp. 1–7.
9. J. Canny, A computational approach to edge detection, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-8**(6), 1986, 679–698.
10. J. Clark and N. Ferrier, Modal control of visual attention, in *Proceedings of the International Conference on Computer Vision, Tarpon Springs, FL, Dec. 1988,* pp. 521–531.
11. L. S. Davis, A survey of edge detection techniques, *Comput. Graphics Image Process.* **4**(3), 1976, 248–270.
12. U. R. Dhond and J. K. Aggarwal, Structure from stereo–A review, *IEEE Trans. Syst. Man Cybernet.* **19**(6), 1989, 1489–1510.
13. W. E. L. Grimson, A computer implementation of a theory of human stereo vision, *Philos. Trans. R. Soc. London B* **292,** 1981, 217–253.
14. M. J. Hannah, Bootstrap stereo, in *Proceedings of ARPA Image Understanding Workshop, College Park, MD, Apr. 1980,* pp. 201–208.
15. W. Hoff and N. Ahuja, Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection, *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(2), 1989, 121–136.
16. M. R. M. Jenkin, J. K. Tsotsos, and G. Dudek, The horoptor and active cyclotorsion, in *Proceedings of IEEE International Conference on Pattern Recognition,* Oct. 1994, pp. A707–710.
17. Y. C. Kim and J. K. Aggarwal, Positioning 3-D objects using stereo images, *IEEE J. Robot. Autom.* **RA-3**(4), 1987, 361–373.
18. D. Marr, *Vision,* W. H. Freeman, San Francisco, 1982.
19. D. Marr and E. Hildreth, Theory of edge detection, *Proc. R. Soc. London* 1980, 187–217.
20. D. Marr, G. Palm, and T. Poggio, Analysis of a cooperative stereo algorithm, *Biol. Cybernet.* **28,** 1978, 223–229.
21. D. Marr and T. Poggio, A computational theory of human stereo vision, *Proc. R. Soc. London B* **204,** 1979, 301–328.
22. L. Matthies and S. A. Shafer, Error modeling in stereo navigation, *IEEE J. Robot. Autom.* **RA-3**(3), 1987, 239–248.
23. G. Medioni and R. Nevatia, Segment-based stereo matching, *Comput. Vision Graphics Image Process.* **31,** 1985, 2–18.
24. H. P. Moravec, Towards automatic visual obstacle avoidance, in *Proceedings of 5th International Joint Conference of Artificial Intelligence, 1977,* p. 584.
25. Y. Ohta and T. Kanade, Stereo by intra- and inter-scanline search using dynamic programming, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-7**(2), 1985, 139–154.
26. T. J. Olson and D. J. Coombs, Real-time vergence control for binocular robots, *Int. J. Comput. Vision* 1991, 881–888.
27. G. Sandini and V. Tagliasco, An anthropomorphic retina-like structure for scene analysis, *Comput. Graphics Image Process.* **14,** 1980, 365–372.
28. J. Weng, N. Ahuja, and T. S. Huang, Matching two perspective views. *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(8), 1992, 806–825.