Qualitative depth from stereo, with applications*

Daphna Weinshall
CBIP, E25-201, MIT, Cambridge MA 02139

Abstract

Obtaining exact depth from binocular disparities is hard if camera calibration is needed. We will show that qualitative information can be obtained from stereo disparities with little computation, and without prior knowledge (or computation) of camera parameters. First, we derive two expressions that order all matched points in the images by depth in two distinct ways from image coordinates only. Using one for tilt estimation and point separation (in depth) demonstrates some anomalies observed in psychophysical experiments, most notably the "induced size effect". We apply the same approach to detect qualitative changes in the curvature of a contour on the surface of an object, with either x- or y-coordinate fixed. Second, we develop an algorithm to compute axes of zero-curvature from disparities alone. The algorithm is shown to be quite robust against violations of its basic assumptions for synthetic data with relatively large controlled deviations. It performs almost as well on real images, as demonstrated on an image of four cans at different orientations.

1 Introduction

Research in early vision regarding stereo seems to be concerned mainly with the correspondence problem, namely, matching points in the left and the right images. Obtaining exact depth values from a stereo pair has been considered a simple exercise whose solution is well known, although it might involve some tedious computations. Thus, it has been implicitly assumed that the final goal of stereo algorithms is to compute an exact depth map using disparity values. The following observations suggest, however, that depth computation from disparity values is not necessarily straightforward or even feasible, and that more qualitative depth information may be more robust and easier to obtain.

First, the depth computation problem reduces to simple trigonometry when the parameters of the cameras, or the eyes, are known. When they are not known, a scheme to compute the camera's parameters from a number of conjugate points (that is, matched pairs of points from the different images) has been devised, involving the solution of a set of nonlinear equations (see for instance [6]). Since the problem has no closed-form solution, and since the data are usually imprecise, a solution is found using iterative methods that minimize squared error. In practice, however, this approach is difficult to implement. The parameters of the cameras must be obtained from data that have errors comparable to the magnitude of the disparity values, which are the raw material used for depth computation (e.g., error due to pixel quantization). In other words, the registration

*Present address: Institute of Computer Science, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel; contact email: daphna@cs.huji.ac.il

problem (namely, finding parameters for the camera's calibration) is much more difficult than the computation of depth from disparity values. Less general methods to perform camera calibration have also been devised, see [14] and [7].

The other observation originates from biological vision. It seems that human vision does not necessarily obtain exact depth values from stereo disparity information alone, see, e.g., [5]. Rather, stereo disparity seems to be used mainly in obtaining qualitative depth information about objects in the field of view. The estimation of the magnitude of this relative depth is possibly dependent on an independent estimation of some physical parameters like the angle of convergence of the eyes.

In view of that, we will show that qualitative relative depth information (ordering) of various kinds can be obtained from only conjugate points in two stereo images easily and reliably, involving almost no computations and independently of the camera's parameters. These orderings will demonstrate some anomalies that are observed in human psychophysics and presently lack other straightforward explanations, most notably the "induced size effect". We will further show that some qualitative shape information can be obtained from image coordinates only. First, one can detect qualitative changes in the curvature of a contour on the surface of any object in the field of view, with either x- or y-coordinate fixed. Similarly, we estimate axes of zero-curvature for objects in the image from disparities alone. This algorithm is tested on synthetic and real data to check robustness against violations of the basic assumptions of the computation and the existence of noise. We then analyze the dependence of errors due to quantization on parameters such as proximity to the axes and the angle of convergence.

2 Basic Geometry

Given two cameras, assume that the principal rays intersect at a fixation point. Also, assume that the epipolar plane of the fixation point (the plane through the principal rays of the cameras, henceforth "base plane") includes the X-axes of both cameras (which are, therefore, epipolar lines by definition). Thus rotation about the principal rays of the cameras is fixed. We will use the following coordinate system (see figure 1): let the fixation point be the origin, the base plane (which passes through this point) be the $X-Z$ plane, and the line perpendicular to this plane through the origin be the Y-axis. On the $X-Z$ plane, the principal rays of both cameras intersect at the origin and create an angle 2μ between them. Let the Z-axis be the angle-bisector of 2μ, and the X-axis perpendicular to the Z-axis. This system is closely related to the Cyclopean coordinate system used in the literature in which the angle-bisector is replaced by the median to the baseline of the cameras and the origin is translated to the midpoint of the baseline. A similar system can be defined for motion if the fixation point is kept constant, that is, the cameras follow a single object. This is more typical of human vision than machine vision.

For a given point P let α denote the angle of tilt and β denote the angle of slant (see figure 1). Thus the Cartesian representation of P is $(\frac{z}{\tan\alpha}, \frac{z}{\tan\beta}, z)$, where z is its depth relative to the fixation point in the above coordinate system. Let (x_l, y_l) and (x_r, y_r) be the Cartesian coordinates of the projection of P on the left and right images respectively (see lower part of figure 1). Using polar coordinates, the two projections can be written as (R_l, ϑ_l) and (R_r, ϑ_r) respectively. Let $\lambda = \frac{\cot\vartheta_r}{\cot\vartheta_l}$. Then the following can be shown to hold (see appendix A):

$$\tan\alpha \quad = \frac{1}{\tan\mu} \cdot \frac{\lambda-1}{\lambda+1} \quad (1)$$

$$\tan\beta \quad = \frac{\cot\vartheta_r - \cot\vartheta_l}{2\sin\mu} \quad (2)$$

Thus, the two angles α and β (of P) depend only on the angle of convergence and the polar angle of the projection (P') on each image. It can be shown that the polar angles are preserved under projection, through any point on the principal ray, onto either a spherical body (like the eye) or a planar one (a camera). There is no dependence on other parameters of the cameras, their relative positions or the angle of gaze ν.

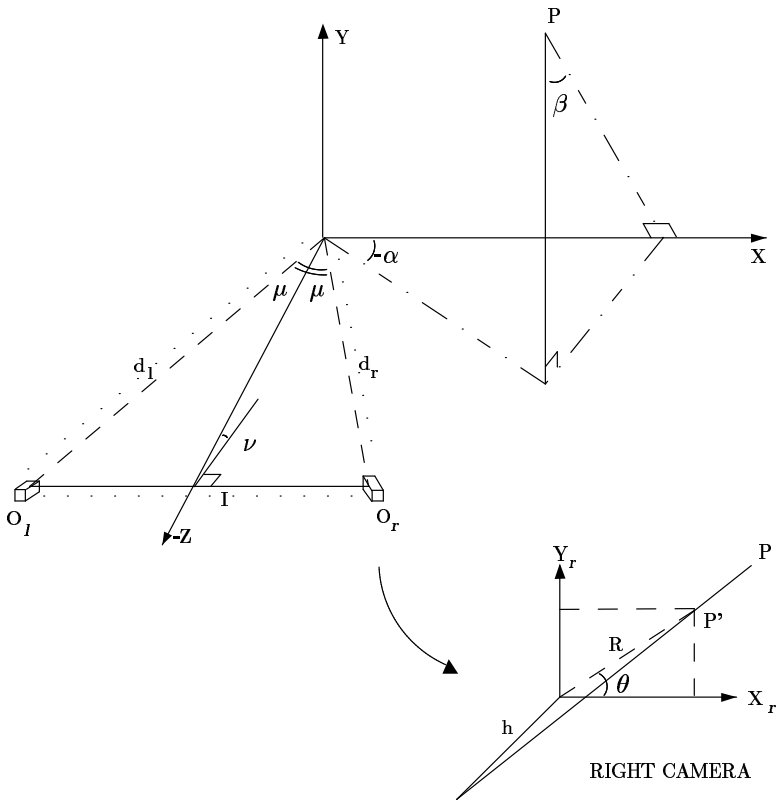

Figure 1: Above, the 3D coordinate system defined by two cameras. Below, the image plane of the right camera.

3 Qualitative Depth

We will use equation (1) of the previous section to obtain an ordering on all matched image points according to their tilt and separately for the left and right halves of the visual field. This ordering is independent of the camera's parameters and demonstrates psychophysical anomalies like the induced effect and others. We will use equation (2) to obtain an expression for the relative depth z. However, this expression will depend on the values of the focal length, the distance between the cameras, and the angle of gaze. An approximate parameter-independent relative depth ordering

will then be obtained from this expression for small angles of convergence 2μ. With additional assumptions and computations, the exact coordinates can be computed (see appendix B).

3.1 Tilt-related Order

From (1) it immediately follows that α is monotonically increasing with λ for a fixed configuration of the cameras. Thus λ defines a mathematical ordering of the matched points in each side of the Y-axis according to their tilt. The ordering defined by λ agrees with the relative depth ordering when comparing points that lie approximately on the same line of sight from the viewer, namely, about the same image x-coordinate.

Note that

$$\ln \lambda = \ln \frac{\cot \vartheta_r}{\cot \vartheta_l} = \ln \frac{x_r/x_l}{y_r/y_l} = (\ln x_r - \ln x_l) - (\ln y_r - \ln y_l) = \Delta(\ln x) - \Delta(\ln y). \tag{3}$$

In other words, if any matching algorithm is applied to the output images of the transformation $T : (x, y) \longrightarrow (\ln x, \ln y)$ performed on the original images, and the disparity vector (Δ_x, Δ_y) is then computed in the usual way, then the difference $\Delta_x - \Delta_y = \ln \lambda$ defines a similar ordering as λ.

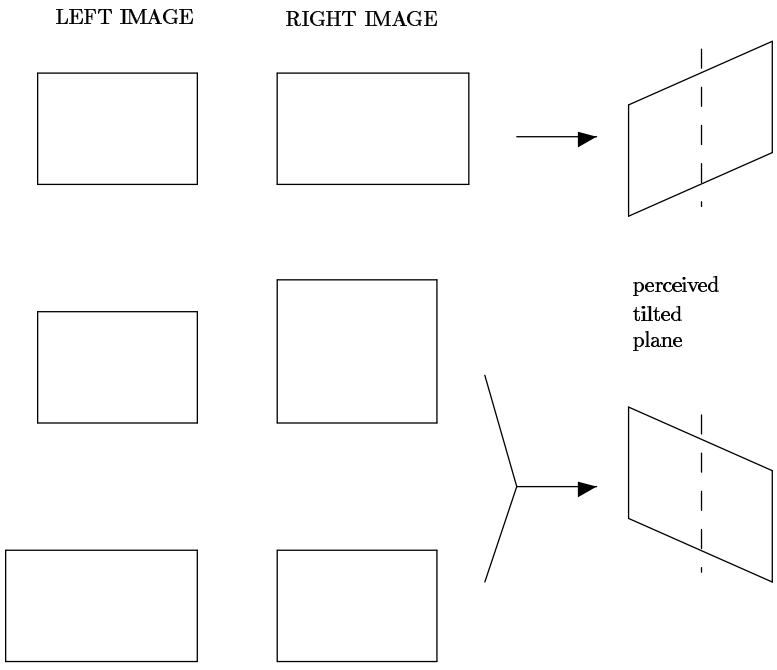

Figure 2: An illustration of the induced-effect. Above - the tilt impression caused by stretching the X-axis of the right image (correct perception). Below - the tilt impression caused by stretching the Y-axis of the right image (wrong perception). The same perception is obtained if the X-axis of the left image is stretched by the same amount.

One prediction of using λ is the "induced effect", the psychophysical effect where a distortion of one image by stretching the Y-axis (the vertical axis) of that image produces a tilt impression

similar to that produced by stretching the X-axis (the horizontal axis) of the other image by the same amount (see figure 2). Whereas the tilt impression caused by stretching the X-axis has a simple geometrical explanation, the reversed tilt impression caused by stretching the Y-axis has none, and has therefore been called an induced effect. Induced, since it is as if the unrealistically magnified Y-axis of one image induces the shrinkage of the X- and Y-axis of that image as a compensation. This effect, first reported by Ogle ([13]), has stimulated extensive research, see [1], [2], [8], [10], [11], and [16].

Estimating tilt by λ gives similar misperception since λ involves only terms of the form $\frac{y}{x}$. Hence multiplying the Y-axis by some number has the same effect as multiplying the X-axis by its inverse. Thus an induced effect is simply a side-effect of using the expression $\frac{\cot \vartheta_r}{\cot \vartheta_l}$. This explanation does not depend on any assumptions and approximations, or the complete recovery of all depth-related parameters of the scene. Other researchers ([10] and [11]) explain the induced effect as a by-product of a specific approximation scheme and a tedious numerical computation; it does not result from an exact solution of the disparity equations. Another computational explanation ([2]) suggests that a distortion occurs in the matching stage, assuming matching is done along horizontal lines only.

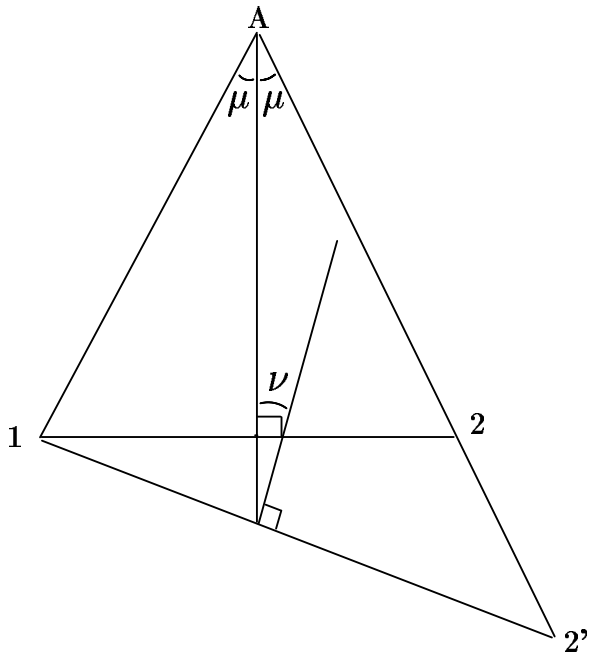

Figure 3: False positive angle of gaze induces the perception of backward motion (point $2'$), whereas the true value 0 shows no such motion (point 2), see text.

Motion also shows an illusion similar to the induced effect (see [16]). In this case, observers reported that a fronto-parallel plane (a plane parallel to the $X - Y$ plane) appeared to be tilting in depth with the right-hand side apparently closer than the left when the monocular image was progressively magnified with head movement to the right and vice versa. λ can account for this phenomenon. Moreover, in this case there is an additional effect - a perceived forward/backward

motion. This could possibly be accounted for by the angle of gaze ν. As will be shown later,

$$\tan\nu \approx \frac{1}{\tan\mu}\cdot\frac{1-\frac{y_r}{y_l}}{1+\frac{y_r}{y_l}}.$$

Thus, a distortion of the y-axis in one image will distort ν (a distortion of the x-axis will affect ν much less). The angle of gaze ν can give the direction of motion as demonstrated in the following example (figure 3): in motion from point 1 to point 2 ν is 0 (the true angle of gaze). Positive ν, the computed angle of gaze, implies motion from 1 to $2'$, that is, backward movement of the head in addition to its left to right movement. Since the head only rotates, the object is perceived as moving backwards.

Quantitatively, since computing λ involves computing ratios of the images x-coordinates and y-coordinates, one should expect computational problems near the X- and Y- axes. It is interesting to note that human performance also deteriorates near the axes, especially near the horizontal one ([1]). This deterioration is accounted for by a smaller probability for the correct detection of the tilt of an oblique line when either the x- or y-axis is magnified, and when the angle between the oblique line and the horizontal axis is around either 0^o or 90^o.

3.2 Depth-related Order

From equation (2) one can obtain an explicit expression for the depth z of a point relative to the fixation point (the origin). First, note that (2) implies

$$z = (\cot\vartheta_r - \cot\vartheta_l)\cdot\frac{y}{2\sin\mu}. \tag{4}$$

Thus, $\chi^y = (\cot\vartheta_r - \cot\vartheta_l)$ defines a relative depth ordering on all the points in space with some constant height y over the base plane. It is shown in appendix A that

$$y = \left\{\frac{I\cos(\mu-\nu)}{\sin 2\mu} - x\sin\mu + z\cos\mu\right\}\cdot\frac{y_r}{h}. \tag{5}$$

Substituting (5) in (4) gives

$$z = \frac{I\cos(\mu-\nu)/\sin 2\mu}{\{\frac{\sin\mu}{x_r-\frac{y_r}{y_l}x_l}[2h+\tan\mu(x_r+\frac{y_r}{y_l}x_l)]-\cos\mu\}}.$$

For an angle of convergence 2μ small enough so that $2h \gg |\tan\mu(x_r+\frac{y_r}{y_l}x_l)|$ we obtain a relative depth ordering on all the points in the visual field by using

$$\chi = x_r - \frac{y_r}{y_l}x_l.$$

Note that if μ is known, the exact depth can be computed up to a scaling factor.

As will be shown in the next section, $\frac{y_r}{y_l} = 1+O(\mu)$. Likewise, since the field of view is bounded by some solid angle $2\xi < 180^o$, it follows that $x < h\tan\xi$. Thus, a sufficient condition for χ to give a correct relative depth ordering is $1 \gg \tan\mu\cdot\tan\xi$. If $2\xi \le 90^o$, which is a generous upper bound for most cameras, it is sufficient if $1 \gg \tan\mu$, or $2\mu \ll 90^o$.

We have obtained a relative depth ordering by an expression close to the x-disparity (corrected for convergence and non zero vergence, that is, nonzero angle of gaze). However, for a fixed convergence angle 2μ, this ordering suffers some distortion compared to the true relative depth ordering, which increases with the horizontal distance from the point of fixation (the x-coordinate).

3.3 Approximation of Vertical Disparities

From the definition of λ and χ it follows that the base plane itself is singular in the sense that these orders are not defined for points on it. One can, however, estimate the orders by substituting $\frac{y_r}{y_l}$ of a matched point far from the base plane. More specifically, for $P = (x, y, z)$ we have

$$\begin{aligned}\frac{y_r}{y_l} &= \frac{d_l}{d_r} + \frac{z}{d_r}\frac{2\sin\mu\tan\nu}{1+\tan\mu\tan\nu} + \frac{x}{d_r}\frac{2\sin\mu}{1+\tan\mu\tan\nu} + o(\tfrac{x}{d_r}, \tfrac{z}{d_r}) \\ &= 1 - \frac{2\tan\mu\tan\nu}{1+\tan\mu\tan\nu} + \frac{z}{d_r}\frac{2\sin\mu\tan\nu}{1+\tan\mu\tan\nu} + \frac{x}{d_r}\frac{2\sin\mu}{1+\tan\mu\tan\nu} + o(\tfrac{x}{d_r}, \tfrac{z}{d_r}) \qquad (1)\end{aligned}$$

Thus, if point P^j is used to approximate point P^i, the error will be:

$$\left(\frac{y_r}{y_l}\right)^i - \left(\frac{y_r}{y_l}\right)^j \approx \frac{2\sin\mu}{1+\tan\mu\tan\nu}\left[\frac{z^i - z^j}{d_r}\tan\nu + \frac{x^i - x^j}{d_r}\right].$$

The error is especially small when the approximating point P^j lies exactly "above" P^i (differs only in the y-coordinate).

One can use as an approximation $\frac{1-\tan\mu\tan\nu}{1+\tan\mu\tan\nu}$ (the first two terms), so that some (possibly independent) estimate of μ (half the angle of convergence) and ν (the angle of gaze) will suffice to assign a rough value to $\frac{y_r}{y_l}$ when no other source of information is available. Note that one cannot take $\frac{y_r}{y_l} \approx 1$ when computing $x_r - \frac{y_r}{y_l}x_l$, as a first order approximation in μ, since $x_r - x_l$ is of the order of magnitude of μ also.

3.4 Support from Psychophysical Results

The orders λ and χ as defined above and the dependency of the scaling coefficients on camera parameters seem to be consistent with the following psychophysical results:

1. Relative depth perception in human vision seems to be more reliable than absolute depth perception. That is, the distinction between different objects at different depth is much more accurate than the estimation of their absolute depth (with no additional information of perspective).

2. The induced effect, as discussed above, is shown to be a side effect of using the tilt-related order λ to estimate the tilt of a plane at the fixation-point. No assumptions on the way the visual system finds and interprets corresponding points is needed. Moreover, this is a "local" explanation of the induced effect in the sense that it allows for opposite induced effects in neighboring spatial regions, in agreement with psychophysical evidence (see [16]). Likewise, this explanation does not imply a perceived asymmetric convergence of the eyes, again in agreement with empirical data. It is interesting to note that λ might be the discrete equivalent to the term of the optical flow field used by Rogers and Koenderink ([16]) to explain the induced effect in motion parallax. Quantitatively, λ is more susceptible to errors near the axes, in agreement with psychophysical experiments ([1]) that show deterioration in human performance of tilt estimation near the axes. Such an effect for a plane not passing through the fixation point will be predicted by the qualitative shape analysis in the next section.

3. Comparison of the depth of features on both sides of the Y-axis is less accurate than if they are on both sides of the X-axis. This is demonstrated in the following experiment ([15]). The first stimulus is a surface whose depth (Z) as a function of X is shown in figure 4a. Here Z is constant with Y. When comparing the depth of the two edges F_l and F_r, that are at the same depth, the observer (wrongly) perceives F_l as closer to her than F_r. In the second stimulus the same configuration is rotated by 90^o so that Z changes with Y as is shown in figure 4b and is constant with X. Now the observer (rightly) perceives F_l and F_r as equidistant. This result is consistent with λ, which orders points on each side of the Y-axis separately. χ is also defined differently on each side of the Y-axis, and it has some distortion as a function of the horizontal distance between the two points.

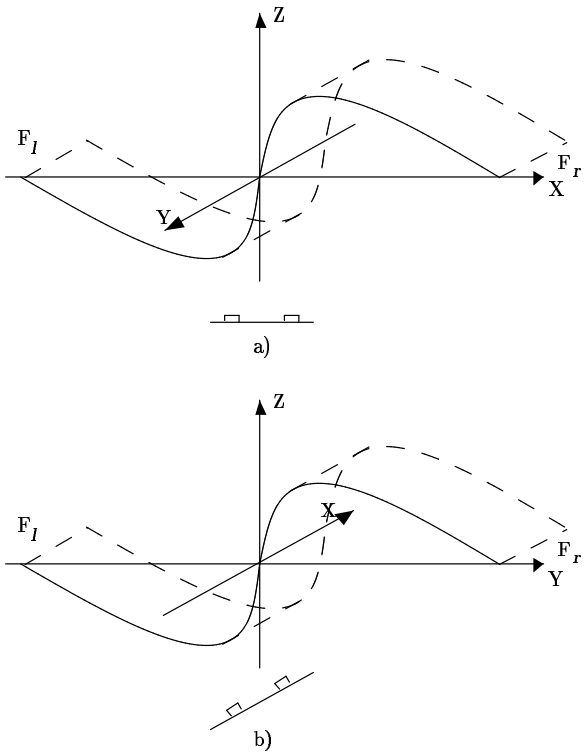

Figure 4: Anisotropy between the horizontal (X) and vertical (Y) dimensions in human depth perception (see text).

4. There is psychophysical evidence for the deterioration in depth discrimination when points are coplanar with the point of fixation (see [12]). Note that λ is constant for points coplanar with the fixation point.

5. There is empirical evidence for the dependence of relative depth perception on external perception of the angle of convergence of the eyes. This is consistent with using the orders λ and χ to evaluate depth with no additional computations.

Moreover, λ and χ^y depend only on the polar angles of the conjugate points in both images, a quantity that is preserved under projection to a spherical body (an eye) or a planar body (a camera). It is interesting to note, in this respect, that the first visual transformation from the retina to $V1$ in primates seems to be in good agreement with the complex-log mapping ([17]), namely: $(x, y) \longrightarrow (\log r, \vartheta)$. This mapping explicitly computes the polar angle ϑ of a point.

4 Qualitative Shape and Applications

We will use the basic relations obtained above to compute some qualitative information on the surface of objects. Robustness against noise and violations of the assumptions of the basic computations will be tested on synthetic and real data.

4.1 Axes of Zero-Curvature

Given three different points on an object P_0, P_1 and P_2, the vector $(P_1 - P_0) \times (P_0 - P_2)$ is parallel to the normal to the plane that passes through the three points. It vanishes if (and only if) P_0, P_1 and P_2 are collinear in space, which is the case for points on a zero-curvature axis (and for only such points) by definition. Moreover, collinear points are projected onto a straight line in each image. Thus, an algorithm to find a zero-curvature axis at a point P_0 on the surface of an object, when a single one exists, is as follows:

Let O_0 be the projection of P_0 on one image plane. For each line in this plane, which passes through O_0 at some angle θ (the lines are parameterized by θ), select two points on line θ in the image plane, denoted by O_1 and O_2. These points are the projections of points P_1 and P_2 that lie on the surface of the same object (this is a condition for the selection of O_1 and O_2). For convenience, select O_1 and O_2 such that each lies on a different side of O_0 on line θ. Next, estimate the expression:

$$\Theta = \mid (P_1 - P_0) \times (P_0 - P_2) \mid .$$

Return the direction θ that minimizes Θ (ideally, under infinite resolution and precision, will be zero).

This is useful when there is one and only one axis of zero-curvature through each point on the object, e.g. cylinders and cones (see [3] for a different approach to this problem). It is most useful for cylinders, where the directions of zero-curvature axes are the same at each point on the object.

Had we the exact depth values at each point, we could compute Θ exactly and find the direction θ that minimizes it. To estimate Θ without knowing the exact depth values, recall that

$$P_i \;=\; z_i(\frac{1}{\tan\alpha_i},\ \frac{1}{\tan\beta_i},\ 1) \;=\; z_i(\tan\mu\frac{\lambda_i+1}{\lambda_i-1},\ 2\sin\mu\frac{1}{\chi_i^y},\ 1)$$

($\lambda = \frac{\cot\vartheta_r}{\cot\vartheta_l}$ and $\chi^y = (\cot\vartheta_r - \cot\vartheta_l)$ as defined above). Let

$$\vec{\Theta} = (P_1 - P_0) \times (P_0 - P_2) = (\Theta_x, \Theta_y, \Theta_z).$$

The following can be readily verified:

$$\Theta_x \qquad = 2z_1 z_2 \sin\mu \cdot \bar{\Theta}_x$$

$$\begin{aligned}\Theta_y &= -z_1 z_2 \tan\mu \cdot \bar{\Theta}_y \\ \Theta_z &= 2 z_1 z_2 \tan\mu \sin\mu \cdot \bar{\Theta}_z\end{aligned} \tag{2}$$

and

$$\begin{aligned}\bar{\Theta}_x &= \tfrac{z_0}{z_2}\left(\tfrac{1}{\chi_1^y} - \tfrac{1}{\chi_0^y}\right) - \left(\tfrac{1}{\chi_1^y} - \tfrac{1}{\chi_2^y}\right) + \tfrac{z_0}{z_1}\left(\tfrac{1}{\chi_0^y} - \tfrac{1}{\chi_2^y}\right) \\ \bar{\Theta}_y &= \tfrac{z_0}{z_2}\left(\tfrac{\lambda_1+1}{\lambda_1-1} - \tfrac{\lambda_0+1}{\lambda_0-1}\right) - \left(\tfrac{\lambda_1+1}{\lambda_1-1} - \tfrac{\lambda_2+1}{\lambda_2-1}\right) + \tfrac{z_0}{z_1}\left(\tfrac{\lambda_0+1}{\lambda_0-1} - \tfrac{\lambda_2+1}{\lambda_2-1}\right) \\ \bar{\Theta}_z &= \tfrac{z_0}{z_2}\left(\tfrac{\lambda_1+1}{\lambda_1-1}\cdot\tfrac{1}{\chi_0^y} - \tfrac{\lambda_0+1}{\lambda_0-1}\cdot\tfrac{1}{\chi_1^y}\right) - \left(\tfrac{\lambda_1+1}{\lambda_1-1}\cdot\tfrac{1}{\chi_2^y} - \tfrac{\lambda_2+1}{\lambda_2-1}\cdot\tfrac{1}{\chi_1^y}\right) \\ &\quad + \tfrac{z_0}{z_1}\left(\tfrac{\lambda_0+1}{\lambda_0-1}\cdot\tfrac{1}{\chi_2^y} - \tfrac{\lambda_2+1}{\lambda_2-1}\cdot\tfrac{1}{\chi_0^y}\right)\end{aligned} \tag{3}$$

For each i, j, λ_i and χ_i^y depend only on the the polar coordinates of the projections of each P_i, and $\frac{z_i}{z_j}$ can be approximated by image coordinates to a first order in the angle of convergence μ. That is,

$$\frac{z_i}{z_j} \approx \frac{\chi_i}{\chi_j}, \quad \chi = x_r - \frac{y_r}{y_l} x_l \ .$$

Thus $\bar{\Theta}_x$, $\bar{\Theta}_y$, and $\bar{\Theta}_z$ can be estimated from image coordinates only.

As noted above, the estimated axis of zero-curvature at point O_0 will be the axis that obtains the minimum of $\Theta = (\Theta_x^2 + \Theta_y^2 + \Theta_z^2)^{\frac{1}{2}}$. However, since in practice Θ will probably rarely obtain 0, it is possible that an unfortunate choice of O_1 and O_2 for some θ's and the fact that Θ is not normalized will lead to a bad estimate. Nevertheless, our algorithm minimizes the approximate expression $\tilde{\Theta} = (2\bar{\Theta}_x)^2 + (\bar{\Theta}_y)^2 + (\frac{1}{2}\bar{\Theta}_z)^2$.

We first tested the algorithm on some images of cans at different orientations in space. In the example of figure 5, the camera was moved manually to obtain a stereo pair. The fixation point (and hence the origin of each image coordinate system) was taken to be the center of the right image and the corresponding point in the left image (which in the above "bad" example is a few pixels to the left of the center of the left image). We still assume that the $X - Z$ plane approximates the base plane (first section). The two 256x256 images have been matched using a parallel motion algorithm implemented on the Connection Machine ([9]), and its output has been smoothed by averaging with a 3x3 window over neighboring pixels. In a fixed region at the center of each object, the direction of the zero-curvature axis has been estimated using the above algorithm at each pixel. The direction obtained by the largest number of pixels in the region was selected as a final estimate. In an image containing four cylindrical objects at various orientations, the true axis of zero-curvature has been obtained for three (figure 5). A rather good approximation has been obtained for the fourth, where the "second best" direction has been selected (we have used a rather coarse quantization of directions). The error in the fourth can may possibly be due to the fact that this can practically lies on the X-axis, where quantization errors seem to prevail (see figure 6). Additional errors may occur if the central region is not chosen appropriately, that is, if it lies too close to the boundary of an object or if it covers area with little texture.

We then used synthetic images of cylinders and cones to test the robustness of the algorithm against errors and deviations from its basic assumptions. Table 1 summarizes the results for synthetic cylinders and cones defined by

$$\frac{(x - x_0)^2}{a^2} + \frac{(z - z_0)^2}{b^2} = 1 \text{ and } \frac{(\mathrm{x} - \mathrm{x}_0)^2}{\mathrm{a}^2} + \frac{(\mathrm{z} - \mathrm{z}_0)^2}{\mathrm{b}^2} = \frac{(\mathrm{y} - \mathrm{y}_0)^2}{\mathrm{c}^2}$$

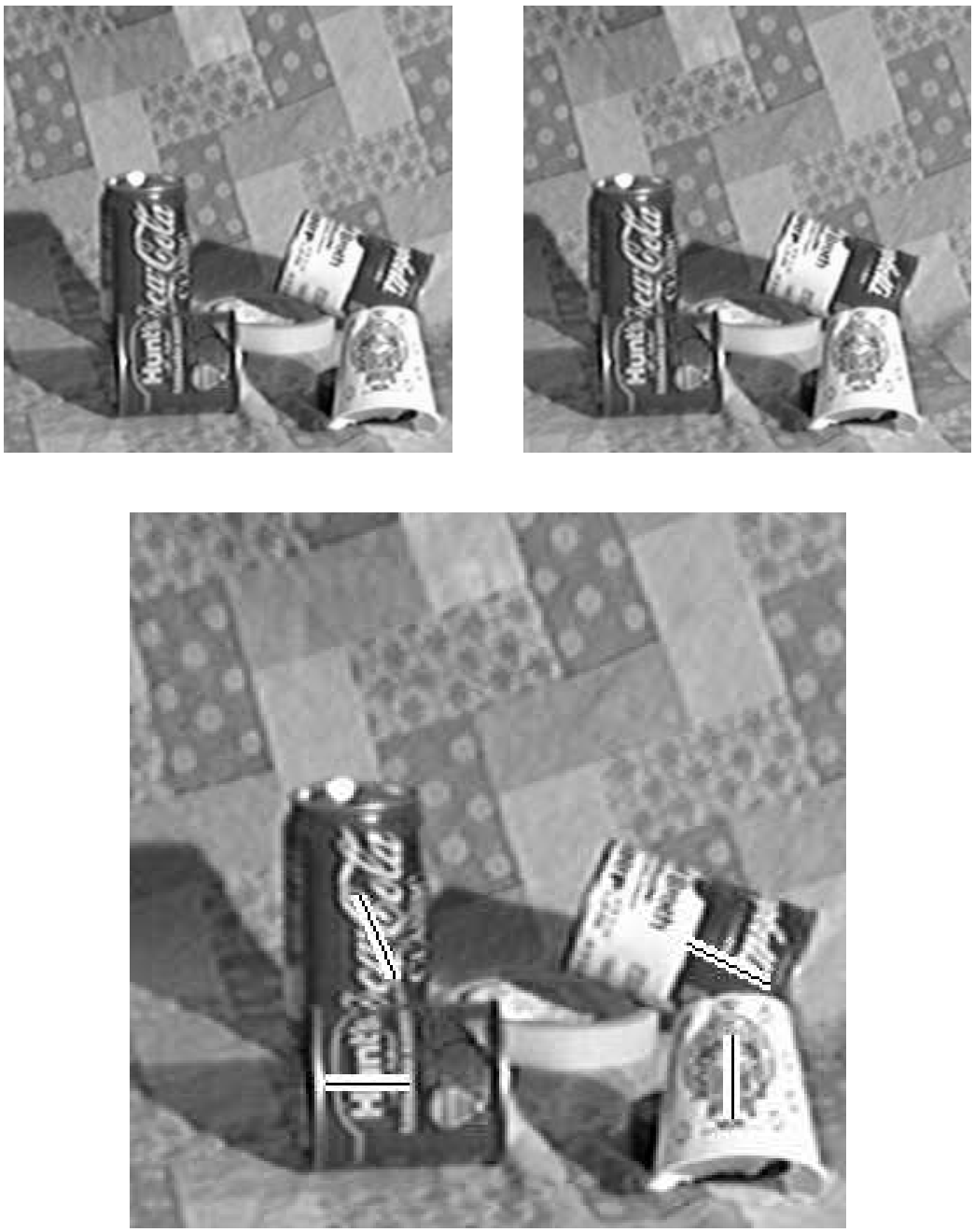

Figure 5: Axes of zero-curvature obtained from "qualitative" shape. Above is the stereo pair. Below is the left image where white lines on each object mark the axes of zero-curvature found by the algorithm.

ζ	ξ	δ_r	δ_l	Δx_r	Δy_r	Δx_l	Δy_l	error
0	0	0	0	0	0	0	0	0
0	0	1	-1	0	0	0	0	9
0	0	5	5	0	0	0	0	7
0	0	0	0	0.1	0.01	-0.01	-0.1	3
0	0	1	1	0.1	0.01	-0.01	-0.1	3
10	30	0	0	0	0	0	0	0
10	30	1	-1	0	0	0	0	1
10	30	5	5	0	0	0	0	2
10	30	0	0	0.1	0.01	-0.01	-0.1	1
10	30	1	1	0.1	0.01	-0.01	-0.1	1
50	0	0	0	0	0	0	0	0
50	0	1	-1	0	0	0	0	0
50	0	5	5	0	0	0	0	1
50	0	0	0	0.1	0.01	-0.01	-0.1	6
50	0	1	1	0.1	0.01	-0.01	-0.1	1
0	90	0	0	0	0	0	0	0
0	90	1	-1	0	0	0	0	8
0	90	5	5	0	0	0	0	4
0	90	0	0	0.1	0.01	-0.01	-0.1	3
0	90	1	1	0.1	0.01	-0.01	-0.1	3
0	0	0	0	0	0	0	0	0
10	30	0	0	0	0	0	0	0
50	0	0	0	0	0	0	0	0
0	90	0	0	0	0	0	0	0

Table 1: Axes of zero-curvature for synthetic cylinders (above the separation line) and cones (below the separation line). See text for the meaning of the different columns. The last column gives the error in degrees (the difference between the true axis of zero-curvature and the one obtained by the above algorithm.)

respectively. In this example we have used $a = 4$, $b = 4$, and $c = 1$. The angle of convergence is 30^o, the distance between the cameras is 13, and the coordinates of the center of the cylinder or cone (x_0, y_0, z_0) are $(50, 50, 20)$ in the 3-D fixation-point coordinate system defined in the first section. However, the algorithm is not sensitive to the exact value of either of these parameters. The point of fixation is equidistant to both cameras.

The objects have been rotated relative to an initial orientation where the main axis is parallel to the y-axis by different values of ξ (rotation about the z-axis) and ζ (rotation about the x-axis), see table 1. We checked robustness against deviations from the basic assumptions of the computation by introducing the following errors: $(\Delta x_i, \Delta y_i)$, $i \epsilon \{r, l\}$ are misalignments of the points of fixation at the two cameras (in image coordinates, where the focal length of the camera is unity); δ_r and δ_l are the angles (in degrees) between the true image X-axis and the X-axis assumed (the line of intersection between the image plane and the base-plane in the right and left images respectively). If any of these errors exists, the X-axis of one image does not lie (exactly) in the base plane, and/or the cameras are not fixated on exactly the same point. Table 1 summarizes the results. The error column gives the difference in degrees between true and estimated axis of zero-curvature in the image plane. (Note that in the image plane a zero-curvature axis is defined by a single angle).

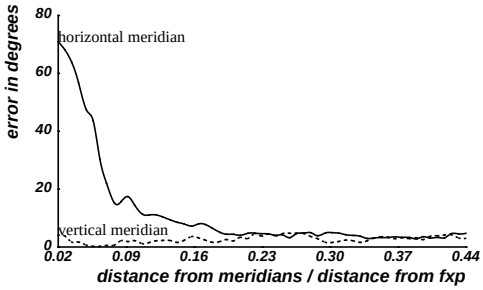

Figure 6: The dependence of quantization errors on proximity to the horizontal and the vertical axes, scaled by the distance from the cameras to the fixation point.

Finally, we measured the dependence of quantization errors on variables such as proximity to the axes, resolution, and the angle of convergence of the cameras. One can see (figure 6) that the performance of the algorithm deteriorates substantially as the reference point on the object approaches the X-axis. There is much less deterioration near the Y-axis. Figure 7 shows the dependence of the error on the angle of convergence of the cameras. Far from the axes, the error is small even for substantial quantization, about a hundred times coarser than human visual hyperacuity (figure 7, below). The functions plotted in figures (6)-(7) measure the average value of the error over some fixed patch of the field of view with only one parameter varying. Some smoothing has been applied to these plots (each point has been averaged with its two nearest neighbors).

4.2 Ordering of Surface Normals

For any two points P_1 and P_2, where $\vec{P}_1 = z_1(\frac{1}{\tan\alpha_1}, \frac{1}{\tan\beta_1}, 1)$ and $\vec{P}_2 = z_2(\frac{1}{\tan\alpha_2}, \frac{1}{\tan\beta_2}, 1)$, let $\vec{N} = \vec{P}_1 \times \vec{P}_2$. $\vec{N}$ is perpendicular to $(\vec{P}_1 - \vec{P}_2)$. (It is actually proportional to the normal to the

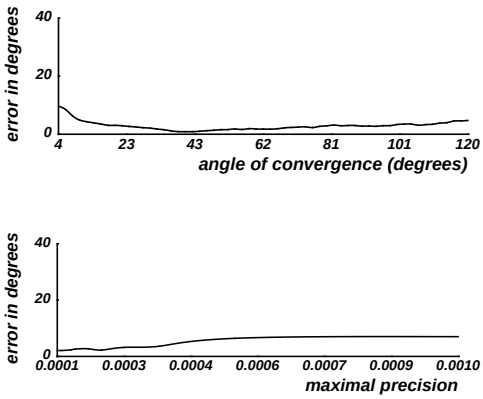

Figure 7: The dependence of quantization error on the angle of convergence of the cameras (above) and the amount of quantization (below).

plane passing through P_1, P_2, and the fixation point.) After some calculations, it can be shown that

$$\begin{aligned}\vec{N} &= z(\tfrac{\cot\beta_1-\cot\beta_2}{\cot\alpha_1\cot\beta_2-\cot\alpha_2\cot\beta_1}\ ,\ \tfrac{\cot\alpha_2-\cot\alpha_1}{\cot\alpha_1\cot\beta_2-\cot\alpha_2\cot\beta_1}\ ,\ 1)\\ &= z(\tfrac{1}{\tan\mu}f(\vartheta_l{}^1,\vartheta_r{}^1,\vartheta_l{}^2,\vartheta_r{}^2)\ ,\ \tfrac{1}{\sin\mu}g(\vartheta_l{}^1,\vartheta_r{}^1,\vartheta_l{}^2,\vartheta_r{}^2)\ ,\ 1)\end{aligned} \tag{4}$$

where

$$\begin{aligned}f(\vartheta_l{}^1,\vartheta_r{}^1,\vartheta_l{}^2,\vartheta_r{}^2) &= \tfrac{(\cot\vartheta_l{}^2-\cot\vartheta_l{}^1)-(\cot\vartheta_r{}^2-\cot\vartheta_r{}^1)}{(\cot\vartheta_r{}^2-\cot\vartheta_r{}^1)+(\cot\vartheta_l{}^2-\cot\vartheta_l{}^1)}\\ g(\vartheta_l{}^1,\vartheta_r{}^1,\vartheta_l{}^2,\vartheta_r{}^2) &= \tfrac{\cot\vartheta_r{}^2\cot\vartheta_l{}^1-\cot\vartheta_l{}^2\cot\vartheta_r{}^1}{(\cot\vartheta_r{}^2-\cot\vartheta_r{}^1)+(\cot\vartheta_l{}^2-\cot\vartheta_l{}^1)}\end{aligned} \tag{5}$$

Thus, as long as $f(\vartheta_l{}^1,\vartheta_r{}^1,\vartheta_l{}^2,\vartheta_r{}^2)$ and $g(\vartheta_l{}^1,\vartheta_r{}^1,\vartheta_l{}^2,\vartheta_r{}^2)$ remain constant, which can be determined from image coordinates only, the points are coplanar (among themselves and with the fixation point), or the object at the fixation-point is planar. Note that λ is obtained from f when $\cot\vartheta_l{}^2 = \cot\vartheta_r{}^2 = 0$ ($g = 0$ then).

For any object it is possible to obtain qualitative information about its surface along any contour, with either x or y fixed. Take a contour on the surface with some fixed y-coordinate, and let P_1 and P_2 be two points on it. Since the y-coordinate of $\vec{P_1}-\vec{P_2}$ is 0, the projection of $\vec{N}$ on the $X-Z$ plane, $\vec{n} = z(\frac{1}{\tan\mu}f(\vartheta_l{}^1,\vartheta_r{}^1,\vartheta_l{}^2,\vartheta_r{}^2)\ ,1)$, is perpendicular to the projection of $\vec{P_1}-\vec{P_2}$. Thus, for fixed y, the one dimensional boundary contour is convex when $f(\vartheta_l{}^1,\vartheta_r{}^1,\vartheta_l{}^2,\vartheta_r{}^2)$ increases with increasing x, concave when $f(\vartheta_l{}^1,\vartheta_r{}^1,\vartheta_l{}^2,\vartheta_r{}^2)$ decreases, and linear when $f(\vartheta_l{}^1,\vartheta_r{}^1,\vartheta_l{}^2,\vartheta_r{}^2)$ remains constant. Note that χ^y can be obtained from f since the sign of f determines relative depth between two points with fixed y-coordinate. The same qualitative description can be obtained for any boundary contour with fixed x from following $g(\vartheta_l{}^1,\vartheta_r{}^1,\vartheta_l{}^2,\vartheta_r{}^2)$ with increasing y. This qualitative description depends only on image coordinates (specifically, on the polar angles of conjugate points). Thus it predicts an "induced effect" for planes that do not include the point of fixation. Obtaining this description is not straightforward, though, since such contours in the 3D coordinate system will be usually mapped to oblique lines in the image plane.

In the general case we estimate the normal to the plane passing through three points P_1, P_2, and P_3 to a first order in μ and the x-disparity

$$\frac{x_r - \frac{y_r}{y_l}x_l}{2h\sin\mu}.$$

In this case, after substituting $K \cdot (x_r - \frac{y_r}{y_l}x_l)$ as an approximation for z, where K is some constant that depends on μ, ν, and h, one gets an expression for the general normal $\vec{N}_G$:

$$\begin{aligned} \vec{N}_G = & \quad (\vec{P}_1 - \vec{P}_2) \times (\vec{P}_2 - \vec{P}_3) \\ \approx & \quad W(V \cdot F(x_r^1, x_l^1, y_r^1, y_l^1, x_r^2, x_l^2, y_r^2, y_l^2, x_r^3, x_l^3, y_r^3, y_l^3), \\ & \quad U \cdot G(x_r^1, x_l^1, y_r^1, y_l^1, x_r^2, x_l^2, y_r^2, y_l^2, x_r^3, x_l^3, y_r^3, y_l^3),\ 1) \end{aligned} \tag{6}$$

where W, V, and U are some constants that depend on μ, ν and h. $F()$ and $G()$ are some functions of images coordinates only. Once again, one can verify approximate planarity of surfaces anywhere in the field of view when $F()$ and $G()$ remain constant.

5 Summary

The goal of this work has been to obtain and use qualitative information from a stereo pair, with as few computations as possible and with a minimal dependence on the camera and scene parameters. First, we have shown that points in a stereo pair, once matched to each other, can be ordered in two related ways according to their tilt (λ) and their depth (χ). These orders are completely determined by image coordinates of conjugate points, and no camera or scene parameters are needed. λ and some variation of χ (χ^y) depend only on the polar angles of the conjugate points in both images, a quantity that is preserved under projection to a spherical body (an eye) or a planar body (a camera). We discussed some psychophysical (and neurophysiological) results that can be understood by the use of such orders. Most notably, the use of λ for tilt estimation predicts "the induced size effect", an unusual behavior of the human visual system that lacks other straightforward explanation.

Some qualitative shape information has then been obtained from stereo disparities. We first developed an algorithm to detect axes of zero-curvature on objects (where a single such axis exists). This algorithm performed well on synthetic data even when the basic assumptions of the computation have been substantially violated. It performed almost as well on real images. In one example of an image of four cans, the true axes of zero-curvature have been found for three cans, and the "second best" axis has been found for the fourth. Finally we showed that one can detect qualitative changes in the curvature of a contour on the surface of any object in the field of view, with either x- or y-coordinate fixed.

A Derivation of Some Geometrical Relations

Figure 8 illustrates the projection of a point in space (P) onto the image plane. From similarity of triangles it follows that

$$\frac{\bar{BP}}{\bar{B'P'}} = \frac{\bar{BO}}{\bar{B'O}} = \frac{\bar{DO}}{\bar{A'O}} = \frac{\bar{AO} - \bar{AB}\cos\varphi}{\bar{A'O}}.$$

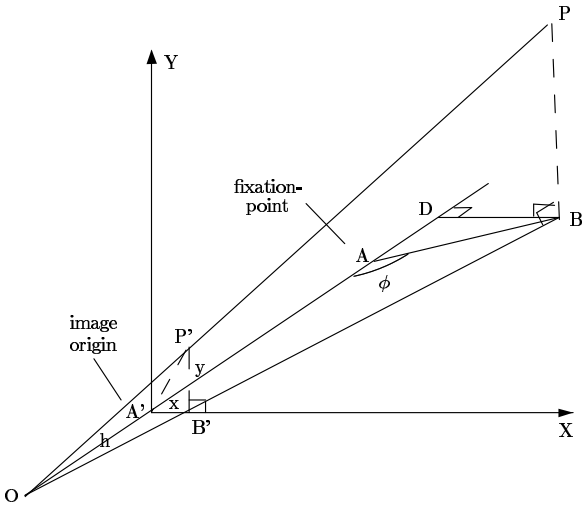

Figure 8: The projection of point P on the image plane.

$B\bar{'}P'$ is the image coordinate y, $A\bar{'}O$ is the focal length h and $A\bar{O}$ is the distance from the fixation point to the camera d. Thus

$$y = \frac{hB\bar{P}}{d - A\bar{B}\cos\varphi}. \tag{6}$$

In a similar way

$$x = \frac{hA\bar{B}\sin\varphi}{d - A\bar{B}\cos\varphi}.$$

Thus

$$\frac{x}{y} = \frac{A\bar{B}}{B\bar{P}} \cdot \sin\varphi.$$

The assumption that the base plane intersects both cameras X-axes implies that the same geometry holds for both cameras in the sense that the segments $B\bar{P}$ and $A\bar{B}$ are identical. We add indices l and r for the variables of the left and right cameras respectively. Then

$$\frac{x_r}{y_r} = \frac{A\bar{B}}{B\bar{P}} \cdot \sin\varphi_r \quad , \quad \frac{x_l}{y_l} = \frac{A\bar{B}}{B\bar{P}} \cdot \sin\varphi_l, \tag{7}$$

and finally

$$\frac{x_r}{y_r} / \frac{x_l}{y_l} = \frac{\sin\varphi_r}{\sin\varphi_l}. \tag{8}$$

!

From figure 9 it follows that

$$\varphi_r = \alpha + 90^o - \mu \quad , \quad \varphi_l = \alpha + 90^o + \mu.$$

Thus

$$\frac{\sin\varphi_r}{\sin\varphi_l} = \frac{1 + \tan\alpha\tan\mu}{1 - \tan\alpha\tan\mu}. \tag{9}$$

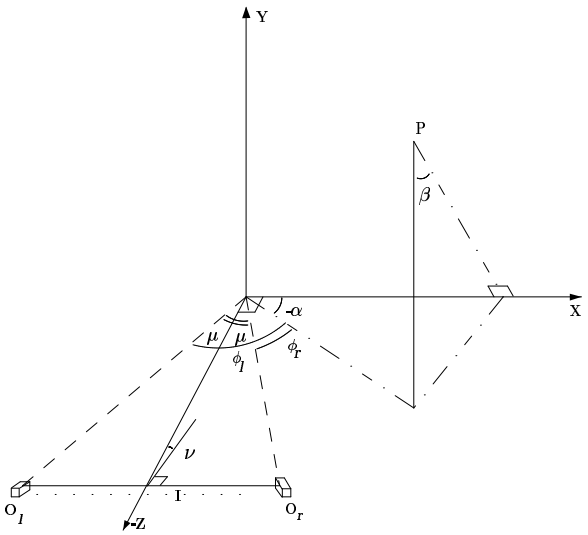

Figure 9: The 3D coordinate system with both cameras.

From (8) and (9) we get

$$\lambda = \frac{x_r}{y_r} / \frac{x_l}{y_l} = \frac{1 + \tan\alpha \tan\mu}{1 - \tan\alpha \tan\mu} \implies \tan\alpha \tan\mu = \frac{\lambda - 1}{\lambda + 1},$$

which gives equation (1).

Similarly,

$$\begin{aligned} \cot\vartheta_r - \cot\vartheta_l &= \tfrac{x_r}{y_r} - \tfrac{x_l}{y_l} = \tfrac{\bar{AB}}{\bar{BP}}(\sin\varphi_r - \sin\varphi_l) \\ &= \tfrac{\bar{AB}}{\bar{BP}} 2\sin\alpha \sin\mu = 2\sin\mu \tfrac{z}{y} \end{aligned} \tag{7}$$

Since by definition $\tan\beta = \frac{z}{y}$, we immediately obtain equation (2).

Finally, considering the right image with no loss of generality, we have from (6),

$$y_r = \frac{h\bar{BP}}{d_r - \bar{AB}\cos\varphi_r} = \frac{hy}{d_r - x\sin\mu + z\cos\mu}.$$

From figure 9 one can see that $d_r = \frac{I\cos(\mu-\nu)}{\sin 2\mu}$, so that

$$y = \left(\frac{I\cos(\mu-\nu)}{\sin 2\mu} - x\sin\mu + z\cos\mu\right) \cdot \frac{y_r}{h}. \tag{5}$$

B Computation of the Exact Depth

We will compute the exact tilt and depth to a first order in the angle of convergence 2μ, following Mayhew & Longuet-Higgins' method (in [11]) to compute tilt and slant of a plane through the

fixation point. The following scheme, however, will be simpler and involve less and more rigorous assumptions (we shall only assume small 2μ as implied above). Since, to a first order in μ, $\tan 2\mu \approx \frac{I\cos(\nu)}{R}$, where R is the distance between the fixation point and the midpoint of the interocular line (the nose), our computations will be correct to a first order in $(\frac{1}{R})$.

Let (x, y) and (x', y') denote the image coordinates of a certain point in space on the two cameras respectively. Let $\hat{\alpha}$ and $\hat{\beta}$ denote the parameters of a plane that passes through a given point in space and the fixation point in the above coordinate system, so that $Z = \hat{\alpha}X + \hat{\beta}Y$. Thus $\hat{\alpha}$ is $\tan(\alpha)$ in the previous notations if $\hat{\beta} = 0$ and vice versa.

Then, to a first order in μ, we have ([4])

$$\begin{aligned} \Delta x = x' - x = \quad & [(\hat{\alpha}\cos(\nu) + \sin(\nu))x + \hat{\beta}\cos(\nu)y \\ & +(\cos(\nu) - \hat{\alpha}\sin(\nu))x^2 - \hat{\beta}\sin(\nu)xy] \cdot I/R, \\ \Delta y = y' - y = \quad & [\sin(\nu)y + (\cos(\nu) - \hat{\alpha}\sin(\nu))xy \\ & -\hat{\beta}\sin(\nu)y^2] \cdot I/R \end{aligned} \tag{8}$$

(The coordinate system used to obtain (10) is the Cyclopean coordinate system. This, however, does not change the results when changing to our coordinate system since the angle-bisector and the median are the same line to a first order in μ and the translation of the origin has been taken into account in the definition of the target plane.)

It is usually possible to consider only the plane passing through a point in space and the fixation point and is perpendicular to the base plane (that is, $\hat{\beta} = 0$). This plane is parametrized only by $\hat{\alpha}$. Thus, we have

$$\begin{aligned} \frac{\Delta y}{y} \quad & = [\sin(\nu) + (\cos(\nu) - \hat{\alpha}\sin(\nu))x] \cdot I/R \\ & = [\tan(\nu) + (1 - \hat{\alpha}\tan(\nu))x] \cdot \tan(2\mu) label11 \end{aligned} \tag{9}$$

Let (x_1 , y_1 , Δx_1 , Δy_1) be the coordinates and disparities of a point on the vertical axis, so that $x_1 \approx 0$. Then we have

$$\frac{\Delta y_1}{y_1} = \tan(\nu) \cdot \tan(2\mu).$$

Let (x_2 , y_2 , Δx_2 , Δy_2) be the coordinates of a point with $\hat{\alpha} \approx 0$. Such a point, if it exists, can be easily identified since it satisfies $\frac{y'}{y} \approx \frac{x'}{x}$. Then we have:

$$x_2 \cdot \tan(2\mu) = \frac{\Delta y_2}{y_2} - \tan(\nu) \cdot \tan(2\mu) = \frac{\Delta y_2}{y_2} - \frac{\Delta y_1}{y_1} = \frac{y_2'}{y_2} - \frac{y_1'}{y_1}.$$

In other words,

$$\tan(2\mu) = \frac{1}{x_2} \cdot [\frac{y_2'}{y_2} - \frac{y_1'}{y_1}].$$

Now, for any point (x, y) in the image we have, using (10) with $\hat{\beta} = 0$:

$$\frac{x'}{x} - \frac{y'}{y} = \frac{\Delta x}{x} - \frac{\Delta y}{y} = \hat{\alpha} \cdot I\cos(\nu)/R = \hat{\alpha} \cdot \tan(2\mu).$$

This leads to the final equations:

$$\tan(2\mu) = \frac{1}{x_2} \cdot [\frac{y_2'}{y_2} - \frac{y_1'}{y_1}], \tag{12}$$

and

$$\hat{\alpha} = \frac{\frac{x'}{x} - \frac{y'}{y}}{\tan(2\mu)} \; ; \quad \tan(\nu) = \frac{\frac{\Delta y_1}{y_1}}{\tan(2\mu)} \; ; \quad R = \frac{I}{\sqrt{\tan^2(2\mu) + \frac{\Delta y_1}{y_1}^2}}.$$

The ratio $\frac{y'}{y}$ near the Y-axis is relatively reliable and easy to obtain. However, a point with $\hat{\alpha} \approx 0$ does not necessarily exist, in which case we can solve the initial scheme directly. The set of equations remained to be solved is

$$\begin{aligned} \tan(2\mu) - \hat{\alpha} \cdot \frac{\Delta y_1{}'}{y_1} &= \frac{1}{x} \cdot (\frac{y'}{y} - \frac{y_1'}{y_1}) \\ \tan(2\mu) \cdot \hat{\alpha} &= \frac{x'}{x} - \frac{y'}{y} \end{aligned} \tag{10}$$

which reduces to a second degree polynomial in $\hat{\alpha}$.

Acknowledgements: I thank Tommy Poggio, Jim Little, Davi Geiger, and Walter Gillett for their help. The psychophysical aspect of this work gained from discussions with Terry Boult and Heinrich Bülthoff. This research was done partly in the MIT AI Laboratory. It was supported by a Chaim Weizmann postdoctoral fellowship, a grant from the Sloan foundation, a grant from the ONR, DARPA and the NSF.

C References

1. A. Arditi, The dependence of the induced effect on orientation and a hypothesis concerning disparity computations in general, *Vision Res.* 22, 1982, 247-256.

2. A. Arditi, L. Kaufman, and J. A. Movshon, A simple explanation of the induced size effect, *Vision Res.* 21, 1981, 755-764.

3. H. Printz, Finding the Orientation of a Cone or Cylinder, *Proc. IEEE Workshop on Computer Vision*, Computer Society of the IEEE, 94-99, 1987.

4. H. C. Longuet-Higgins, and K. Prazdny, The interpretation of a moving retinal image, *Proc. R. Soc. Lond. B* 208, 385-397, 1980.

5. J. M. Foley, and W. Richards, Effects of voluntary eye movement and convergence on the binocular appreciation of depth, *Perception & psychophysics* 11(6), 1972, 423-427.

6. B. K. P. Horn, *Robot Vision*, MIT Press & McGraw-Hill, 1986.

7. H. C. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections, *Nature* 293, 1981, 133-135.

8. H. C. Longuet-Higgins, The role of vertical dimension in stereoscopic vision, *Perception* 11, 1982, 377-386.

9. J. Little, H. Bülthoff, and T. Poggio, Parallel Optical Flow Computation, *Proc. Image Understanding Workshop*, L. Bauman (ed.), Science Applications International Corp., McLean, VA, 915-920, 1987.

10. J. E. W. Mayhew, The interpretation of stereo-disparity information: the computation of surface orientation and depth, *Perception* 11, 1982, 387-403.

11. J. E. W. Mayhew, and H. C. Longuet-Higgins, A computational model of binocular depth perception, *Nature* 297 (3), 1982, 376-378.

12. G. J. Mitchison, and G. Westheimer, The perception of depth in simple figures, *Vision Res.* 24, No-9, 1984, 1063-1073.

13. K. N. Ogle, Induced size effect. I. A new phenomenon in binocular space perception associated with the relative sizes of the images of the two eyes, *Archs Ophthal* 20, 1938, 604-623.

14. K. Prazdny, Stereoscopic matching, eye position, and absolute depth, *Flair Draft*, 1981.

15. B. J. Rogers, and M. E. Graham, Anisotropies in the perception of three-dimensional surfaces, *Science* 221, 1983, 1409-1411.

16. B. J. Rogers, and J. Koenderink, Monocular aniseikonia: a motion parallax analogue of the disparity-induced effect, *Nature* 322(3), 1986, 62-63.

17. E. L. Schwartz, Anatomical and physiological correlates of visual computation from striate to infero-temporal cortex, *IEEE Trans. on Systems, Man, and Cybernetics* SMC-14 (2), 1984, 257-271.