Determining Three-Dimensional Shape from Orientation and Spatial Frequency Disparities I — Using Corresponding Line Elements

David G. Jones[†] Jitendra Malik[‡]

Technical Report UCB–CSD 91–656

October, 1991

Abstract

When a surface slanted away from the fronto-parallel plane is viewed binocularly, surface markings and texture are imaged with slightly different orientations and degrees of foreshortening. These orientation and spatial frequency disparities are systematically related to surface slant and tilt and could potentially be exploited by biological and machine vision systems. There is evidence suggesting that human stereopsis has a mechanism that specifically makes use of orientation and spatial frequency disparities, in addition to the usual cue of horizontal positional disparity.

In this paper we derive constraint equations relating orientation and spatial frequency disparities to the local surface normal. We derive necessary and sufficient conditions for recovering surface normals: (i) Two measurements of orientation disparity, or (ii) One measurement of orientation disparity and associated spatial frequency disparity. These conditions are readily met in local regions of real images, for example in texture patches and in the neighborhood of brightness edges and lines that are curved or form corners and junctions. We develop a least squares algorithm that provides more reliable computation of 3-D surface normals when more than the minimum number of orientation and spatial frequency disparities are available. Experimental results are presented to demonstrate the success of this approach.

[†]McGill University
Department of Electrical Engineering
Montréal, Québec, Canada H3A 2A7
phone: (514) 398-8348 fax: 398-7348
email: djones@lightning.mcrcim.mcgill.edu

[‡]University of California, Berkeley
Computer Science Division
Berkeley, California, USA 94720
phone: (510) 642-7597 fax: 642-5775
email: jmalik@robotics.berkeley.edu

This work has been supported by a grant to DJ from the Natural Sciences and Engineering Research Council of Canada (OGP0105912) and by a National Science Foundation PYI award (IRI-8957274) to JM.

1 Introduction

Stereopsis has traditionally been viewed as a source of depth information. When we view a three-dimensional scene with our two eyes, the small positional differences of corresponding points in the two images give information about the relative distances to those points in the scene. Viewing geometry, when it is known, provides the calibration function relating disparity to absolute depth. To describe three-dimensional shape, the surface normal, $\boldsymbol{n}(x, y)$, can then be computed by differentiating the interpolated surface $z(x, y)$ (Grimson, 1981).

However, there are other cues available under binocular viewing that can provide direct information about surface orientation. When we view a surface that is not fronto-parallel, surface markings or textures will be imaged with slightly different orientations and degrees of foreshortening in the two views (Fig. 1). These orientation and spatial frequency disparities are systematically related to the local three-dimensional surface orientation. Psychophysicists have demonstrated (see Section 2 of this paper) that humans are able to exploit these cues to determine surface orientation, even when positional disparity information is absent or inconsistent.

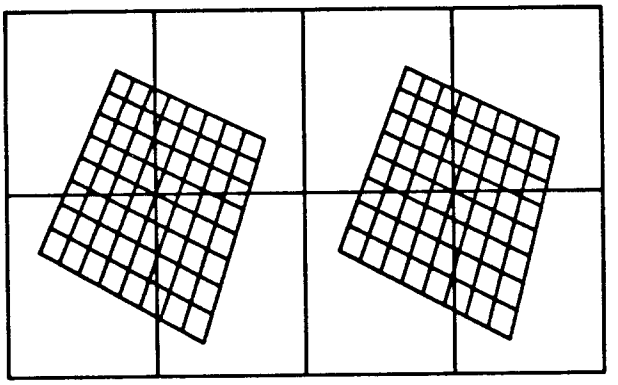


Figure 1: Stereo pair of a planar surface tilted in depth. A careful comparison of the two views reveals a slightly different orientation and spacing for corresponding grid lines drawn on the surface.

There has been very little work investigating the use of these cues in computational vision. In fact, it is quite common in computational stereo vision to simply ignore the orientation and spatial frequency differences, or image distortions, that occur when viewing surfaces tilted in depth. These differences are then a source of error in computational schemes which try to find matches on the assumption that corresponding patches (or edges) must be identical or very nearly so.

Some approaches acknowledge the existence of these image distortions, but still treat them as noise to be tolerated as opposed to an additional signal that may exploited (Arnold and Binford, 1980; Kass, 1983). For example Kass (1987), in a framework based on matching filter outputs, calculates bounds on the expected range of filter output differences that can arise because of the image distortions when viewing slanted surfaces and uses this to guide the selection of a match criterion. The match criterion or threshold can be selected to balance the tradeoff between false-positive matches and correct matches that are improperly ruled out by the criterion. It is very much in the spirit of *coping with* and not *using* image distortions.

A few approaches seek to cope with image distortions in an iterative framework. These methods typically start off with the initial assumption that disparity is locally constant, and then from initial estimates of positional disparity, guess at the parameters of the image distortion and locally

transform the image to compensate so that image regions can be compared once again with the assumption that corresponding regions are merely translated copies of one another (Mori et al., 1973; Quam, 1984; Witkin et al., 1987). The intent is that this procedure will converge, but the initial estimates of positional disparity, which may be made with quite inappropriate assumptions, are relied upon to guide the convergence. In the case where image regions considered might be as large as 64 x 64 pixels, this repeated "warping" of the input image regions can be quite a costly computation. As a model of human stereopsis, this seems an unlikely mechanism, especially since it is unclear whether the "images" are even available at the earliest stages of cortical processing. Having a mechanism so specific to stereopsis that warps and re-filters the visual input seems a heavy price to pay, especially in the presence of the myriad other aspects of human vision.

This paper and its companion paper (Jones and Malik, 1991c) develop novel computational methods for directly recovering surface orientation by exploiting these orientation and spatial disparity cues. The scheme developed in this paper is based on directly exploiting the constraint equations relating orientation and spatial frequency disparities to the local surface normal, while the companion paper is based on recovering H_x, H_y — two parameters which characterize the local affine transformation from one view to the other. Our work was done in the context of a filtering model for computational stereopsis (Jones and Malik, 1990; Jones and Malik, 1991a; Jones and Malik, 1991b) where the outputs of a set of linear filters at a point are used for matching. However, the algorithms we have developed can also be utilized in conjunction with edge-based or area-correlation based approaches.

The only significant previous computational models of how orientation and spatial frequency disparity may be exploited are due to Koenderink and Van Doorn (1976) and Wildes (1991). These models are based on an affine approximation of the transformation from one view to the other which is then decomposed into expansion, rotation, and deformation components. For the special case of orientation disparity, Wildes (1991) shows how surface orientation can be recovered from measurements on three nearby pairs of corresponding line elements (Canny edges).

Our approach has the following advantages:

1. Treatment of both orientation and spatial frequency disparities.
2. The benefit of making use of all the data. While measurements on three pairs may be adequate in principle, using minimal information leads to much greater susceptibility to noise.

This paper is organized as follows: Section 2 reviews relevant psychophysical evidence for the importance of orientation and spatial frequency disparity. Section 3 develops the relationship between surface geometry and orientation and spatial frequency disparity. In section 4, the objective is to develop intution about the magnitude and range of these diaparities. In section 5 a least squares algorithm for utilizing this information is developed and evaluated.

2 Human Visual Psychophysics

Explicit awareness of the cues of orientation and spatial frequency disparity and evidence that they might be important is a relatively recent development in the study of human vision. Some evidence is outlined below.

2.1 Spatial Frequency Disparity

When viewing a visual stimulus in which the left and right eyes see vertical sinusoidal gratings with different spatial frequencies, the perception is that of a planar surface rotated about the vertical axis (Blakemore, 1970). When the spatial frequency is higher for the right eye, the surface appears to slant away from the observer to the left. The correspondence between bars in the grating, however, provides a positional disparity cue, leaving it unclear whether the spatial frequency difference itself is utilized. To eliminate this correspondence, the grating can be replaced by uncorrelated dynamic visual noise, filtered to contain a certain spatial frequency band — giving the appearance of a mixture of blurred vertical bars of different widths, rapidly and randomly changing. When there is a spatial frequency difference, but no systematic positional correspondence, the perception of slant remains (Tyler and Sutter, 1979).

2.2 Orientation Disparity

A similar stimulus can be constructed to test whether, in the absence of systematic positional correspondence, an orientation difference in the two eyes is sufficient to lead the perception of a surface tilted in depth. Using uncorrelated dynamic random lines, with a slightly different orientation in each eye, there is also a consistent perception of slant (von der Heydt et al., 1981).

In much the same way that random dot stereograms confirmed the existence of a mechanism that makes use of horizontal disparities (Julesz, 1960), these experiments provide strong evidence that the human visual system possesses a mechanism that can and does make use of spatial frequency and orientation disparities in the two retinal images to aid in the perception of surface slant.

2.3 Improved Thresholds from Orientation Disparity

An important question, though, is whether under normal viewing conditions these cues make any difference. It might be argued that horizontal disparities alone are sufficient for the recovery of three-dimensional shape. After all, a surface tilted in depth will give rise to a gradient in the disparity map, so given the output of a horizontal disparity mechanism, the surface orientation could, in principle, be recovered by taking partial derivatives in the estimates of horizontal disparity. In practice, any inaccuracies present in the horizontal disparity estimates will be compounded by taking derivatives. If there were a mechanism that recovered surface shape more directly, by making use of orientation and spatial frequency disparities, then the perception of surface slant would be more accurate.

This question has been addressed experimentally (Rogers and Cagenello, 1989). Consider the task of discerning the direction that a planar disc, ruled with a grid pattern similar to that in Figure 1, is tilted around the vertical axis. Depending on the orientation of the grid lines, there will be differing amounts of orientation disparity. Vertical and horizontal grid lines (0°/90°) give rise to *no* orientation disparity, whereas grid lines at ±45°, give rise to orientation disparities that increase as the surface is rotated about the vertical axis. Equations describing this relationship are described later. On the other hand, grid line intersections provide very good features for establishing positional correspondence and measuring horizontal disparity, regardless of the orientation of the grid. If only horizontal disparities are important in judging depth and surface orientation, then

the smallest noticeable tilt away from fronto-parallel (the psychophysical threshold) should be unaffected by the orientation of the grid pattern on the disc. This is not the case. The just noticeable tilt is over twice as large when there are no orientation disparities (0°/90°) as compared to when there are orientation disparities (±45°).

This evidence supports the idea that even in normal circumstances, orientation disparities, and possibly spatial frequency disparities as well, are used in order to provide a greater accuracy in the perception of three-dimensional shape than could be provided by positional disparities alone.

3 Geometrical Basis for Using Orientation and Spatial Frequency Disparities

This section outlines the geometrical relationships that link the three-dimensional orientation of a surface to the resulting orientation and spatial frequency disparities that are observed in a pair of images. Parameters that can be used to specify three-dimensional surface orientation are described and then it is shown how orientation and spatial frequency disparities depend on these. The task of stereo vision, of course, is to attempt to solve the inverse problem, recovering the surface orientation from the measured disparities in a pair of images. It is shown that either of the following are sufficient to allow, in principle, the recovery of a surface's orientation: a pair of orientation disparities; an orientation disparity and a spatial frequency disparity.

3.1 Coordinates and Parameters

In order to discuss the geometry involved in how the three-dimensional orientation of a surface is related to orientation and spatial frequency disparities a coordinate frame and set of parameters must be established. Consider a small surface patch with some arbitrary texture on it. Without loss of generality, we may consider the appearance of a series of evenly spaced parallel lines on a plane. The results obtained will apply when considering orientation and spatial frequencies of general texture patterns.

Let the fixation point lie at the origin of an object-centered coordinate system, as shown in Figure 2. The x-axis is to the right, the y-axis is up, and the z-axis points towards the viewer. The viewer's eyes (or cameras) lie in the xz-plane and their optical axes make angles $\pm\Delta\phi_y$ with the z-axis, and are not rotated about their optical axes.

To describe the parameters of an arbitrarily oriented plane with a series of evenly spaced parallel lines on it, start with a unit vector pointing along the x-axis, $(1,0,0)$. A rotation ϕ_z around the z-axis allows the pattern to have any orientation on the surface. A rotation of ϕ_x around the x-axis followed by a rotation ϕ_y around the y-axis combine to allow any orientation of the surface itself.

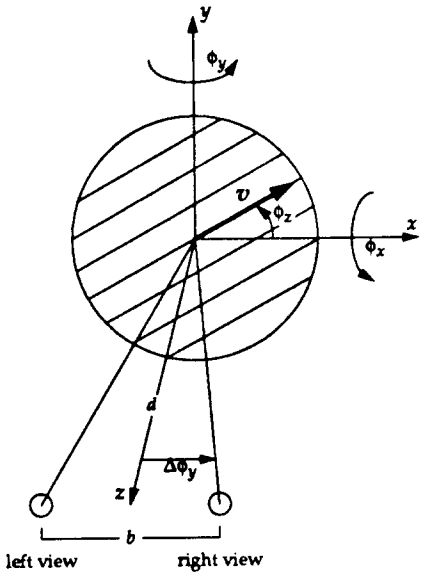


Figure 2: Parameters for specifying the three-dimensional orientation of a small planar surface patch. Parameters for specifying the three-dimensional orientation of a small planar surface patch. A planar surface (*disc*) is viewed at a distance d, from two vantage points separated by a distance b. A three-dimensional vector, $\boldsymbol{v}$, is used as a reference in the direction of a generic surface texture (a set of parallel lines). An arbitrary configuration can be achieved by first rotating the surface pattern ϕ_z around the z-axis, then rotating the surface ϕ_x around the x-axis, and lastly rotating the surface ϕ_y around the y-axis. The different viewpoints can be handled conveniently by adding an additional rotation $\pm\Delta\phi_y$ around the y-axis, where $\Delta\phi = \tan^{-1}(b/2d)$ around the y-axis.

3.2 Projection onto the Image Planes

The three-dimensional vector $\boldsymbol{v}$ resulting from the above transformations can be written concisely:

$$\begin{bmatrix} \\ \boldsymbol{v} \\ \\ \end{bmatrix} = \begin{bmatrix} \cos\phi_y & 0 & \sin\phi_y \\ 0 & 1 & 0 \\ -\sin\phi_y & 0 & \cos\phi_y \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi_x & -\sin\phi_x \\ 0 & \sin\phi_x & \cos\phi_x \end{bmatrix} \begin{bmatrix} \cos\phi_z & -\sin\phi_z & 0 \\ \sin\phi_z & \cos\phi_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} \sin\phi_x \sin\phi_y \sin\phi_z + \cos\phi_y \cos\phi_z \\ \cos\phi_x \sin\phi_z \\ \sin\phi_x \cos\phi_y \sin\phi_z - \sin\phi_y \cos\phi_z \end{bmatrix}$$

The vector $\boldsymbol{v}$ indicates the three-dimensional orientation of the lines ruled on the surface. In order to consider orientation and spatial frequency disparities, this vector must be projected onto the left and right image planes. In orthographic projection, points are projected onto the image along lines normal to the image plane, instead of converging on a focal point. In what follows, orthographic projection will be used, since it provides a very close approximation to perspective projection for the small surface patches under consideration. For orientation disparity, the results are unchanged whether an orthographic or perspective projection is used. For spatial frequency disparity, the difference is negligible under most realistic viewing circumstances (when line spacing is small relative to the viewing distance).

Instead of working with lengths such as the viewing distance d and the baseline separation between the viewpoints b, it will be more convenient to work with an angle $\Delta\phi_y = \tan^{-1}(b/2d)$. Orthographic projection of $\boldsymbol{v}$ can be achieved by replacing ϕ_y with $\phi_y + \Delta\phi_y$ and then discarding the z component to give the two-dimensional image vector $\boldsymbol{v}_l$, the orthographic projection of $\boldsymbol{v}$ onto the left image plane. Similarly, replacing ϕ_y with $\phi_y - \Delta\phi_y$ gives $\boldsymbol{v}_r$, the projection of $\boldsymbol{v}$ on the right image plane.

3.3 Orientation Disparity

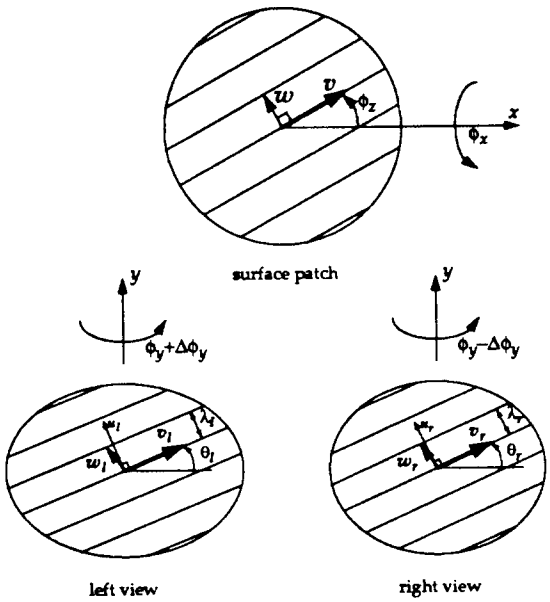


Figure 3: The slightly different appearance of a tilted surface from two viewpoints. A three-dimensional surface patch (*top*), ruled with parallel lines at an orientation given by ϕ_z, is seen from two viewpoints, giving two different images (*bottom*). For reference, three-dimensional vectors lie parallel ($\boldsymbol{v}$) and perpendicular ($\boldsymbol{w}$) to these lines. The surface orientation can be specified by a rotation by ϕ_x around the x-axis, followed by a rotation by ϕ_y around the y-axis. The resulting two-dimensional image textures (*lines*) can be described by their orientation, θ, and spacing, λ. The text describes how orientation disparity, $\theta_r - \theta_l$, and spatial frequency disparity, $\lambda_r - \lambda_l / \frac{1}{2}(\lambda_r + \lambda_l)$, are related to surface orientation ϕ_x, ϕ_y.

Let θ_l and θ_r be the angles the image vectors $\boldsymbol{v}_l$ and $\boldsymbol{v}_r$ make with the x-axis (Fig. 3). These orientations can be easily expressed in terms of the components of the image vectors.

$$\tan\theta_l = \frac{\cos\phi_x \tan\phi_z}{\sin\phi_x \sin(\phi_y + \Delta\phi_y)\tan\phi_z + \cos(\phi_y + \Delta\phi_y)}$$

$$\tan\theta_r = \frac{\cos\phi_x \tan\phi_z}{\sin\phi_x \sin(\phi_y - \Delta\phi_y)\tan\phi_z + \cos(\phi_y - \Delta\phi_y)}$$

This is all that is needed if the goal is to determine the *orientation disparity* $\theta_r - \theta_l$, from a known pattern orientation ϕ_z, a known surface orientation ϕ_x, ϕ_y, and a known view angle $\Delta\phi_y$. This will be the case in Section 4.1 where the probability distribution of orientation disparities is considered.

For now, consider the reverse problem of determining the surface orientation ϕ_x, ϕ_y from an observed orientation disparity. One observation of orientation disparity is insufficient to allow recovery of a surface's orientation. A pair of observations, however, are sufficient as the following will show. Each of the above equations can be rearranged to isolate $\tan\phi_z$.

$$\tan\phi_z = \frac{\cos(\phi_y + \Delta\phi_y)\tan\theta_l}{\cos\phi_x - \sin\phi_x \sin(\phi_y + \Delta\phi_y)\tan\theta_l} = \frac{\cos(\phi_y - \Delta\phi_y)\tan\theta_r}{\cos\phi_x - \sin\phi_x \sin(\phi_y - \Delta\phi_y)\tan\theta_r}$$

Setting these two expressions for $\tan\phi_z$ equal eliminates ϕ_z. The resulting expression can be rearranged to isolate $\tan\phi_x$.

$$\tan\phi_x = \frac{\cos(\phi_y - \Delta\phi_y)\tan\theta_r - \cos(\phi_y + \Delta\phi_y)\tan\theta_l}{\sin(2\Delta\phi_y)\tan\theta_r \tan\theta_l}$$

The orientations of a pair of corresponding line elements in the surface texture, (θ_l, θ_r and θ'_l, θ'_r) give two expressions for $\tan\phi_x$. Setting these equal yields the following expression with ϕ_x eliminated.

$$\frac{\cos(\phi_y + \Delta\phi_y)}{\cos(\phi_y - \Delta\phi_y)} = \frac{\tan\theta_l - \tan\theta'_l}{\tan\theta_r - \tan\theta'_r}$$

This can be simplified by defining $\tau_l = \tan\theta_l - \tan\theta'_l$ and $\tau_r = \tan\theta_r - \tan\theta'_r$, and then rearranged to isolate $\tan\phi_y$.

$$\frac{1 - \tan\phi_y \tan\Delta\phi_y}{1 + \tan\phi_y \tan\Delta\phi_y} = \frac{\tau_l}{\tau_r}$$

$$\tan\phi_y = \frac{1}{\tan\Delta\phi_y} \cdot \frac{\tau_r - \tau_l}{\tau_r + \tau_l}$$

It is now clear that from the orientations $\theta_l, \theta_r, \theta'_l, \theta'_r$ of a pair of corresponding line elements, the three-dimensional orientation of the surface can be recovered by solving first for ϕ_y and then ϕ_x. For completeness, the true orientation ϕ_z, ϕ'_z of each of the line elements on the surface can also be recovered.

As a concrete example, if there is a texture element resembling a '+' on a surface, then the orientation of the surface can be estimated from the orientations of the two line elements observed in two different views. The accuracy of this estimate is naturally limited by the accuracy of the orientation estimates. Furthermore, if the line elements are close to being parallel, then the surface orientation estimate would be less reliable, since the above expressions would be numerically unstable. Any implementation that relied on this relationship between orientation disparity and surface slant would likely be able to provide increased accuracy by making use of several orientation disparity estimates (see Section 5.1).

3.4 Spatial Frequency Disparity

Let λ_l, λ_r be the spacing, and $f_l = 1/\lambda_l$, $f_r = 1/\lambda_r$ be the spatial frequency of the lines in the left and right images (Fig. 3). The following shows how the *spatial frequency disparity* f_r/f_l depends on the texture orientation ϕ_z, the surface orientation ϕ_x, ϕ_y, and the relative viewing distance $\Delta\phi_y$.

Spatial frequency is measured perpendicular to the lines in the image. For this reason, a new unit vector $\boldsymbol{w}$ is introduced which is perpendicular to $\boldsymbol{v}$. This vector indicates the spacing between the lines. An expression for $\boldsymbol{w}$ can be easily obtained from the expression for $\boldsymbol{v}$ by replacing ϕ_z with $\phi_z + 90°$.

$$\boldsymbol{w} = \begin{bmatrix} \sin\phi_x \sin\phi_y \cos\phi_z - \cos\phi_y \sin\phi_z \\ \cos\phi_x \cos\phi_z \\ \sin\phi_x \cos\phi_y \cos\phi_z + \sin\phi_y \sin\phi_z \end{bmatrix}$$

When these three-dimensional vectors, $\boldsymbol{v}$ and $\boldsymbol{w}$, are projected onto an image plane, they generally do not remain perpendicular. In the left view of Figure 3, for example, $\boldsymbol{v}_l$ and $\boldsymbol{w}_l$ are no longer perpendicular. If we let $\boldsymbol{v}_l^\perp = (-\boldsymbol{v}_{ly}, \boldsymbol{v}_{lx})$, then $\boldsymbol{u}_l = \boldsymbol{v}_l^\perp / \|\boldsymbol{v}_l\|$ is a unit vector perpendicular to $\boldsymbol{v}_l$. The length of the component of $\boldsymbol{w}_l$ parallel parallel to $\boldsymbol{u}_l$ is equal to λ_l, the line spacing in the left image.

$$\lambda_l = \frac{\boldsymbol{w}_l \cdot \boldsymbol{v}_l^\perp}{\|\boldsymbol{v}_l\|}$$

Substituting expressions for $\boldsymbol{v}_l$ and $\boldsymbol{w}_l$ gives an expression for the numerator.

$$\begin{aligned} \boldsymbol{w}_l \cdot \boldsymbol{v}_l^\perp &= [\sin\phi_x \sin\phi(\phi_y + \Delta\phi_y)\cos\phi_z - \cos\phi(\phi_y + \Delta\phi_y)\sin\phi_z][-\cos\phi_x \sin\phi_z] \\ &\quad + [\cos\phi_x \cos\phi_z][\sin\phi_x \sin(\phi_y + \Delta\phi_y)\sin\phi_z + \cos(\phi_y + \Delta\phi_y)\cos\phi_z] \\ &= \cos\phi_x \cos(\phi_y + \Delta\phi_y) \end{aligned}$$

A simple expression for the denominator can be found in terms of θ_l, the angle $\boldsymbol{v}_l$ makes with the x-axis.

$$\begin{aligned} \|\boldsymbol{v}_l\| &= \sqrt{\boldsymbol{v}_{lx}^2 + \boldsymbol{v}_{ly}^2} = \sqrt{\boldsymbol{v}_{ly}^2 \left(\frac{\sqrt{\boldsymbol{v}_{lx}^2 + \boldsymbol{v}_{ly}^2}}{\boldsymbol{v}_{ly}} \right)^2} \\ &= \left| \frac{\cos\phi_x \sin\phi_z}{\sin\theta_l} \right| \end{aligned}$$

The same can be done to get an expression for the λ_r, the line spacing in the right image. Combining these gives a concise expression for spatial frequency disparity.

$$\begin{aligned} \frac{f_r}{f_l} = \frac{\lambda_l}{\lambda_r} &= \frac{\boldsymbol{w}_l \cdot \boldsymbol{v}_l^\perp}{\|\boldsymbol{v}_l\|} \frac{\|\boldsymbol{v}_r\|}{\boldsymbol{w}_r \cdot \boldsymbol{v}_r^\perp} \\ &= \left| \frac{\cos(\phi_y + \Delta\phi_y)\sin\theta_l}{\cos(\phi_y - \Delta\phi_y)\sin\theta_r} \right| \end{aligned}$$

Taking the absolute value is actually unnecessary since $\phi_y \pm \Delta\phi_y$ is always in $(-\pi, \pi)$ for the surface to be visible from both viewpoints, and $\sin\theta_l, \sin\theta_r$ always have the same sign.

Once again, if the goal is to determine spatial frequency disparity from a given pattern orientation ϕ_z, surface orientation ϕ_x, ϕ_y, and relative viewing distance $\Delta\phi_y$ then this equation and the previous

ones to determine θ_l, θ_r are all that are needed. This will be used later in Section 4.1 when the distribution of spatial frequency disparities is considered. In that case, it will be more convenient to deal with the spatial frequency difference relative to the mean spatial frequency $\bar{\lambda} = (\lambda_r + \lambda_l)/2$ since, unlike the ratio λ_l/λ_r, this is symmetric about zero spatial frequency disparity.

$$\frac{\lambda_r - \lambda_l}{\bar{\lambda}} = 2\frac{\cos(\phi_y - \Delta\phi_y)\sin\theta_r - \cos(\phi_y + \Delta\phi_y)\sin\theta_l}{\cos(\phi_y - \Delta\phi_y)\sin\theta_r + \cos(\phi_y + \Delta\phi_y)\sin\theta_l}$$

For now, consider again the problem of determining surface orientation. This time, it will be shown that a single observation of orientation disparity and spatial frequency disparity is sufficient. The first expression for spatial frequency disparity can be rearranged to isolate $\tan\phi_y$. Defining $\sigma_r = \lambda_r \sin\theta_l$ and $\sigma_l = \lambda_l \sin\theta_r$ puts this equation in a form that bears a striking resemblance to the previous one for $\tan\phi_y$.

$$\frac{\lambda_l \sin\theta_r}{\lambda_r \sin\theta_l} = \frac{1 - \tan\phi_y \tan\Delta\phi_y}{1 + \tan\phi_y \tan\Delta\phi_y}$$

$$\tan\phi_y = \frac{1}{\tan\Delta\phi_y} \cdot \frac{\sigma_r - \sigma_l}{\sigma_r + \sigma_l}$$

This makes is clear that from the spacing λ_l, λ_r and orientation θ_l, θ_r of a pair of parallel line elements in the surface texture, the three-dimensional orientation of the surface can be recovered, by first solving for ϕ_y and then using the equations from the previous section to solve for ϕ_x and ϕ_z.

As an example, if there is a texture element similar to '||' (or '▮') on a surface, then the surface orientation can be estimated from the orientation and spacing (or width) observed in two different views.

4 Magnitude of Orientation and Spatial Frequency Disparities

Before proceeding, the reader might want to get a feel for these expressions for orientation and spatial frequency by trying particular values for the various parameters. For example, as the viewing distance becomes very large, or equivalently, as the separation between viewpoints becomes very small, then $\Delta\phi_y \to 0$. In the limit, $\theta_r - \theta_l \to 0$ and $(\lambda_r - \lambda_l)/\bar{\lambda} \to 0$, giving no orientation or spatial frequency disparity. The reader can also confirm that when the surface patch is only rotated about the vertical axis ($\phi_x = 0$), horizontal and vertical line elements in the image ($\theta_l = 0°$ or $90°$) have no orientation disparity. (cf. Section 2.3).

In addition to knowing the disparities in specific viewing situations, it is also important to know how large they are in general, especially if we plan to measure them. Visualizing this information all at once in a single graph is difficult, since even when the viewing distance $\Delta\phi_y$ is fixed, there are four dimensions of interest, namely, the texture orientation ϕ_z, the surface orientation ϕ_x, ϕ_y, and the resulting disparity. In the 3-D graphs below, orientation and spatial frequencies that arise in typical human viewing conditions are shown as a function of texture orientation ϕ_z and one parameter of surface tilt (ϕ_x or ϕ_y). A reasonable choice for "typical" human viewing conditions is taken to be a baseline of 7 cm and a viewing distance of 1 metre, giving $\Delta\phi_y = 2.0°$.

The magnitude of the orientation disparity, $|\theta_r - \theta_l|$, arising when a surface is rotated around either the horizontal or vertical axis, is shown in Figure 4. When a surface is rotated around the horizontal axis, horizontal surface markings give no orientation disparity but orientations closer to vertical give progressively larger orientation disparities, with disparities for near-vertical surface markings being quite large. When a surface is rotated around the vertical axis, surface markings that are precisely horizontal or vertical give no orientation disparity, but surface markings at other orientations (especially just off horizontal) can give rise to rather large disparities, easily exceeding 5°. Given that the human visual system can resolve differences in orientation as small as 1/3° (Watt, 1984), even surfaces tilted a relatively small amount away from fronto-parallel still give readily detectable orientation disparities.

The magnitude of the spatial frequency disparity, $|\lambda_r - \lambda_l/\bar{\lambda}|$, arising when a surface is rotated around either the horizontal or vertical axis is shown in Figure 5. When a surface is rotated around the horizontal axis, surface markings that are precisely horizontal or vertical give rise to no spatial frequency disparity, but surface markings at other orientations (especially just off vertical) can give rise to rather large disparities, easily exceeding 10%. When a surface is rotated around the vertical axis, spatial frequency disparity is always present except for surface markings that are precisely horizontal. Since the human visual system can resolve differences in spatial frequency as small as 2% (Mayer and Kim, 1986), the disparities arising from viewing surfaces tilted in depth are well within the discriminable range.

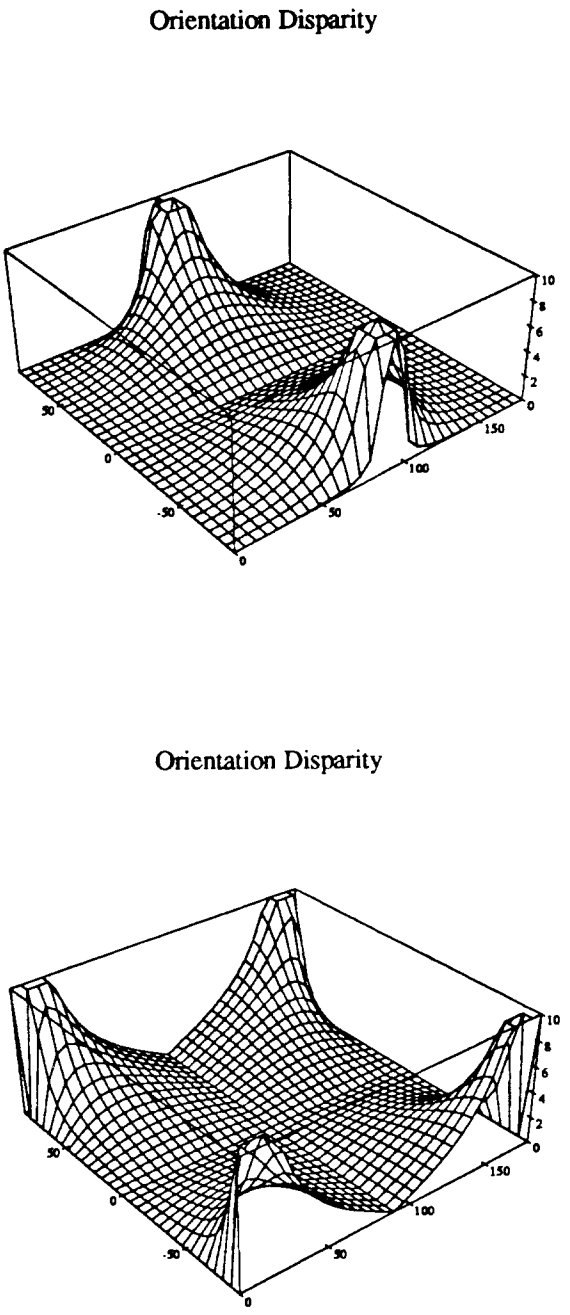
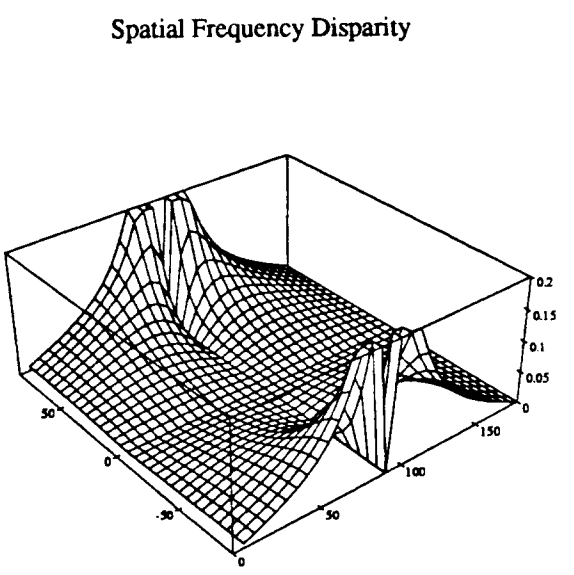


Figure 4: Orientation disparities arising when viewing a surface rotated about either the x-axis or y-axis. The orientation of an element of surface texture is specified by an angle ϕ_x ($0°, 180°$ horizontal; $90°$ vertical). The graphs plot as height, the orientation disparity, in degrees, resulting when the surface being viewed has been rotated either ϕ_x around the x-axis (A), or ϕ_y around the y-axis (B). A surface is fronto-parallel when both ϕ_x and ϕ_y are zero.



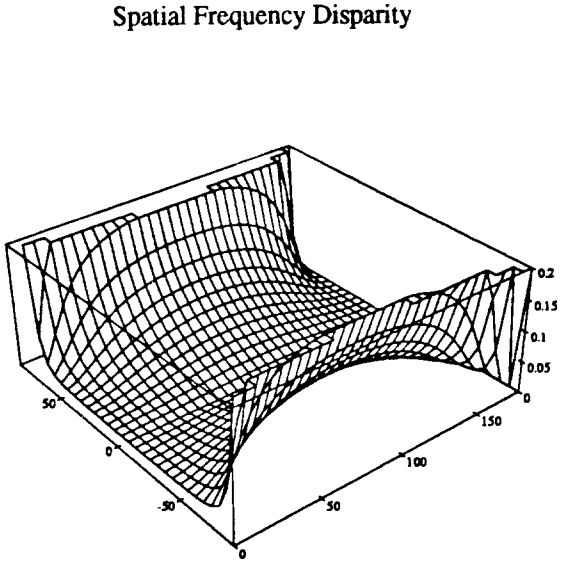


Figure 5: Spatial frequency disparities arising when viewing a surface rotated about either the x-axis or y-axis. As in Fig. 4, the orientation of an element of surface texture is specified by ϕ_x. The graphs plot as height, the spatial frequency disparity, in percent, resulting when the surface being viewed has been rotated either ϕ_x around the x-axis (A), or ϕ_y around the y-axis (B).

4.1 Expected Disparity Ranges

Orientation and spatial frequency disparities are systematically related to how a surface is tilted in depth. The goal is to make use of this information to accurately infer the local three-dimensional surface orientation, or in general, the surface shape. For computational approaches to stereo vision that *do not* try to make of use of this information, these orientation and spatial frequency disparities are a source of *noise*. At best then, these other approaches can be designed to *tolerate* the fact that corresponding image patches contain features or edges with dissimilar orientations and dissimilar lengths or spacings. From this standpoint, it is important to know, on average, how dissimilar corresponding image features can be.

For line segments ("edges"), it is possible to analytically determine the *expected range* of orientation and length differences if all three-dimensional orientations of this line segment are equally probable. For typical human viewing conditions, the probability distribution of orientation differences has a half-width at half-height of 3°, and the probability distribution of length differences has a half-width at half-height of 7% (Arnold and Binford, 1980). These values give some measure of the range of edge orientation and length differences that need to be *tolerated* when matching *edges*.

The goal here, however, is not to ignore these differences, but rather to make use of them to infer how a surface is tilted in depth. There are two aspects of the analysis mentioned above that are unsatisfactory. First, that analysis is in terms of individual line segments and not surface patches of solid objects. Though this distinction is sometimes ignored (Kass, 1987), it is important because a surface can be oriented in such a way that opposite sides of the surface are visible from each viewpoint, or in the case of a solid object, the surface is only visible from one viewpoint because of self-occlusion. Second, in designing a scheme for making use of these differences, it would be nice to have more complete information about the expected ranges of these differences. For example, if a scheme can only cope with some range of orientation differences, then that range could be chosen to include, say 90% or 95% of all possible orientation differences.

Instead of a long analytic derivation involving probabilities, it is more intuitively understandable, though no less accurate, to perform a Monte Carlo simulation to determine experimentally the probability distributions of orientation and spatial frequency disparity. The following procedure was followed to perform this Monte Carlo simulation. First, a pattern orientation ϕ_z is randomly selected from the range $[0°, 180°)$, with all orientations having equal probability. Second, a unit surface normal $\vec{n}$ is randomly selected so that all orientations are equally probable. (This is equivalent to saying $\vec{n}$ is a uniformly random point on the surface of the Gaussian sphere). From this, it is straightforward to calculate the parameters of surface orientation, ϕ_x and ϕ_y. If the front of the resulting surface patch is not visible from both viewpoints, it is discarded, and the process is repeated. Given these parameters, plus some fixed viewing angle $\Delta\phi_y$, it is straightforward to calculate (and store) the resulting orientation and spatial frequency disparities. After many such trials, the experimental probability distributions accurately reflect the theoretical probability distributions. For the results presented here, one million trials were used.

The probability distribution of orientation disparity for two representative viewing conditions is shown in Fig. 6A. As noted earlier, orientation disparity depends on the viewing angle, $2\Delta\phi_y$. Two noteworthy values of $\Delta\phi_y$ are 2.0°, which corresponds to a human with eye separation 7 cm viewing a tilted surface at a distance of 1 metre, and 19.3°, which corresponds to two successive aerial photographs taken at a separation of 700 metres at a height of 1 kilometre.

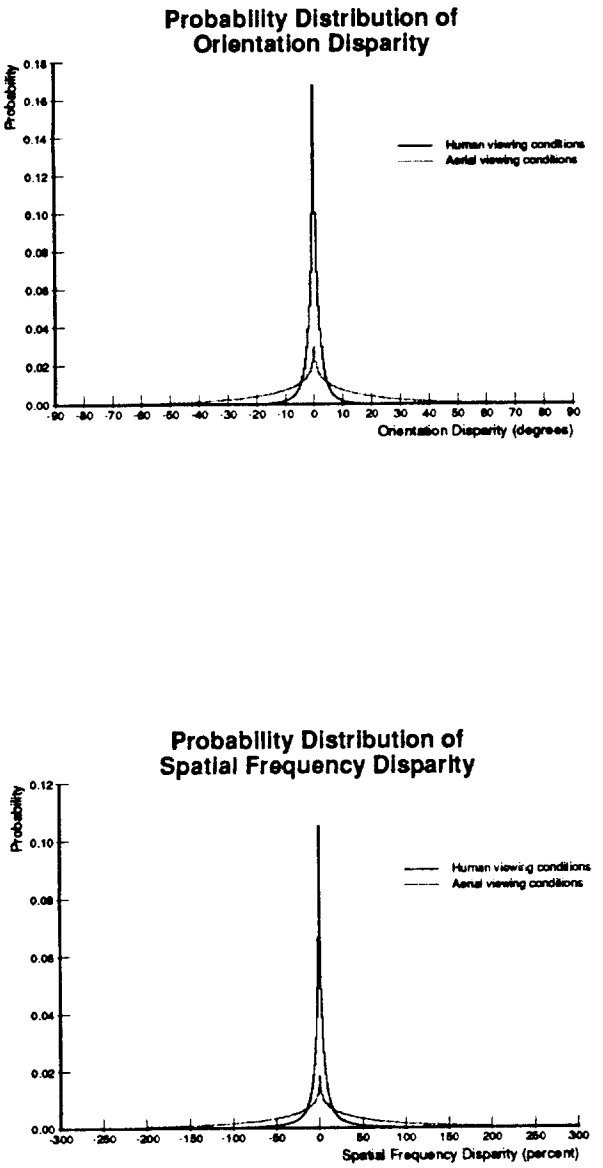


Figure 6: Probability distribution of orientation and spatial frequency disparities. Assuming a uniform random distribution of surface orientations, the probability distributions for orientation disparity (A) and spatial frequency disparity (B) are plotted, following a Monte Carlo simulation with 1,000,000 trials. Distributions are shown for two different representative stereo baselines. Human viewing conditions (*solid line*) assume an eye separation of 7 centimetres and a viewing distance of 1 metre. Aerial viewing conditions (*dotted line*) assume a camera separation of 700 metres and a viewing distance of 1 kilometre.

The distributions are symmetric, so it is known *a priori* that the median should be exactly zero. The median orientation disparity found experimentally was 0.0007 for human viewing conditions and 0.0022 for aerial viewing conditions. This gives a good indication of the accuracy of the results and the number of significant digits in values determined by the simulation.

Given these probability distributions, it is possible to make assertions such as: For a human at a viewing distance of 1 metre, 90% of the time, the orientation disparity is less than 5.5°. The table below gives the range of orientation disparity for various percentage intervals centered at the median, and two viewing conditions.

Orientation Disparity

percent	human vision $\Delta\phi_y = 2.0°$	aerial photography $\Delta\phi_y = 19.3°$
25	± 0.43°	± 3.76°
50	± 1.23°	±10.79°
75	± 2.82°	±24.37°
90	± 5.47°	±44.90°
95	± 8.12°	±62.62°

The probability distribution of spatial frequency disparity for two viewing conditions is shown in Figure 6B. Once again, these distributions are symmetric and the medians of 0.0023 and −0.022 for human and aerial viewing conditions give an indication of the accuracy of the method. The table below gives the range of spatial frequency disparity for various percentage intervals centered at the median, and for two viewing conditions.

Spatial Frequency Disparity

percent	human vision $\Delta\phi_y = 2.0°$	aerial photography $\Delta\phi_y = 19.3°$
25	± 1.61%	± 12.81%
50	± 4.93%	± 36.06%
75	±13.50%	± 79.62%
90	±34.37%	±132.35%
95	±60.03%	±160.99%

If there were some reason to believe that not all surface orientations are equally probable, the above Monte Carlo simulations could easily be modified to take this into account. For example, for a given surface area, the image area subtended is smaller for surfaces slanted further away from fronto-parallel. This foreshortening effect is proportional to the cosine of the slant angle. It might be argued that this should be taken into account when considering the distribution of surface orientations seen in an image. Another argument could be made that in certain man-made environments, surface orientations are quite non-uniformly distributed, with vertical and horizontal surfaces being much more common, though it is their orientation with respect to the viewer that is important.

5 Recovering Surface Orientation from Corresponding Line Elements

5.1 Reliable Estimation of Surface Orientation

While it has been demonstrated that a pair of orientation disparities or an orientation and spatial frequency disparity pair are mathematically sufficient to recover surface orientation, in practice it would be desirable to be able to combine several disparities to arrive at an estimate of surface orientation that would be reliable in the presence of measurement errors, noise, or outliers. The class of methods that immediately come to mind is "least-squares".

Using the equations developed earlier, and given n observations of corresponding orientations, $\theta_{r_i}, \theta_{l_i}$, and possibly spatial periods, $\lambda_{r_i}, \lambda_{l_i}$, the goal is to reliably estimate surface orientation, ϕ_x, ϕ_y. This may at first glance seem like a non-trivial undertaking since the equation relating surface orientation to orientation disparity (Section 3.3) or orientation and spatial frequency disparities (Section 3.4) are non-linear. This is a concern since it could mean that a costly iterative procedure might be needed to compute the least-squares estimate of surface orientation. The key, however, is that all the nonlinearities are confined to values that are either measured θ_{r_i}, θ_{l_i}, or assumed to be a (possibly known) constant $\Delta\phi_y$. This is demonstrated by using a trigonometric identity to rewrite the equations as follows.

$$\begin{aligned}\tan\phi_x &= \frac{\cos\phi_y\cos(\Delta\phi_y)(\tan\theta_{r_i}-\tan\theta_{l_i})+\sin\phi_y\sin(\Delta\phi_y)(\tan\theta_{r_i}+\tan\theta_{l_i})}{\sin(2\Delta\phi_y)\tan\theta_{r_i}\tan\theta_{l_i}} \\ &= a_i\cos\phi_y + b_i\sin\phi_y\end{aligned}$$

All the measured or observed quantities can be collapsed into two easily computed numbers a_i, b_i. If $\boldsymbol{e}_{\phi_y}$ is defined to be the unit vector $(\cos\phi_y, \sin\phi_y)$, then this gives a series of equations that can be written in the form

$$\tan\phi_x = (a_i, b_i)\cdot\boldsymbol{e}_{\phi_y}$$

This has a straightforward geometric interpretation depicted in Figure 7. For a given surface orientation, all the observations (a_i, b_i) should lie along a straight line. The angle made by the perpendicular to that straight line gives ϕ_y and the distance along that perpendicular from the origin to the line is $\tan\phi_x$. Estimating ϕ_x, ϕ_y is equivalent estimating the best-fitting straight line through the points (a_i, b_i). Since any two points specify a line, it is obvious why two error-free points are adequate. The best-fitting line will go through the centroid, so subtracting the centroid from each point would allow ϕ_y to be estimated first, from the best-fitting line through the origin, and then ϕ_x can be determined second.

The results are much the same when several spatial frequency disparities are also available. The key equation relating orientation and spatial frequency to surface orientation can be rewritten

$$\frac{\lambda_l\sin\theta_r}{\lambda_r\sin\theta_l} = \frac{\cos\phi_y\cos(\Delta\phi_y)-\sin\phi_y\sin(\Delta\phi_y)}{\cos\phi_y\cos(\Delta\phi_y)+\sin\phi_y\sin(\Delta\phi_y)}$$

$$\begin{aligned}0 &= \cos\phi_y(\lambda_l\sin\theta_r-\lambda_r\sin\theta_l)+\sin\phi_y(\lambda_l\sin\theta_r+\lambda_r\sin\theta_l) \\ 0 &= a_i\cos\phi_y + b_i\sin\phi_y\end{aligned}$$

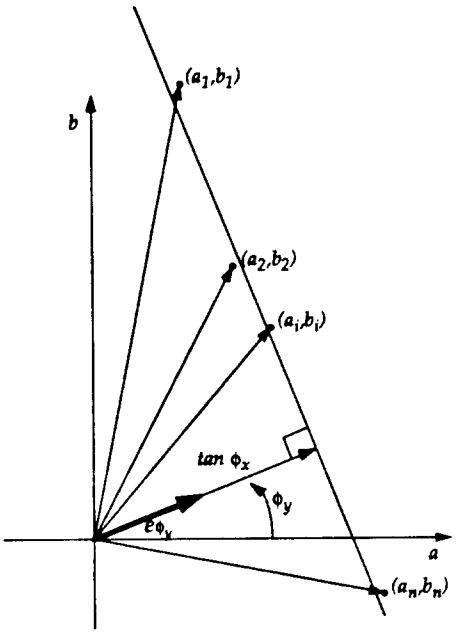


Figure 7: Determining surface orientation from several observations of orientation disparity. From the orientations of corresponding line elements lying on a surface, the data points (a_i, b_i) are easily calculated. These points must lie along a straight line. A unit vector e_{ϕ_y} perpendicular to the best-fitting line gives one parameter of surface orientation, ϕ_y. The projection of the vectors (a_i, b_i) on this unit vector, or the distance of the best-fitting line from the origin, is $\tan\phi_x$, giving the other parameter of surface orientation.

This problem also reduces to fitting a line to the observed data points (a_i, b_i), but this time the line is constrained to go through the origin. The direction perpendicular to this line gives ϕ_y, and again, ϕ_x can be calculated given the value of ϕ_y.

For either case, any procedure for fitting a line to a set of points can be employed. The important thing to note, however, is that it is possible to calculate the parameters ϕ_x, ϕ_y of surface orientation *directly*, without an iterative algorithm.

5.2 Experimental Evaluation

The computer implementation of this method was evaluated using test stereo pairs such as those shown in Figure 8. Each of these nine stereo pairs depicts a textured, planar surface with known surface orientation. The center surface is fronto-parallel, the middle row is rotated around the vertical axis, the middle column is rotated around the horizontal axis, and the four corners are rotated around both axes. The white squares marked on these surfaces are to help make the horizontal compression/expansion and the vertical skew more salient. The image distortion between the two views in a stereo pair can be characterized by two parameters, H_x and H_y, corresponding to these two effects. This characterization is associated with an alternative approach to utilizing orientation and spatial frequency disparities and is discussed in a companion paper (Jones and Malik, 1991c).

In total, 25 test surface orientations were used, with H_x and H_y taking on values: $0.0, \pm 0.1, \pm 0.2$. For each test surface orientation, 50 indepently generated stereo pairs of a randomly textured

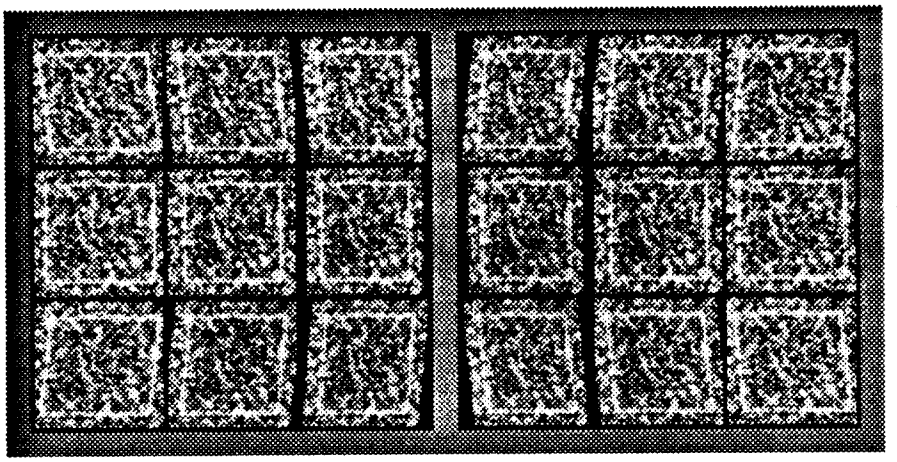


Figure 8: Stereo pair of a planar surfaces tilted in depth.

plane were created. On each trial, the predominant local orientation at the center of the textured surface was determined for both the left and right views. To do this, the image was convolved with rotated copies of an elongated Gaussian derivative filter $G'_{\sigma_1}(x)G_{\sigma_2}(y)$ with $\sigma_2 : \sigma_1$ set at $3 : 1$. The orientation associated with the maximum response of this family (found by parabolic interpolation) was used as the estimate of local orientation. Tests showed that this scheme was able to provide orientation estimates accurate to about 1°. Our previous experience with running edge detectors on real images is consistent with this — orientation estimates accurate to within 1° can be reasonably obtained but one may not expect higher precision in general.

In the presence of this orientation uncertainty, we expect to have trouble for nearly horizontal edges ($\theta \approx 0$). It is easy to see this from the equations for a_i, b_i below, where $1/\tan\theta$ terms cause numerical difficulties.

$$a_i = \frac{1}{2\sin(\Delta\phi_y)}\left(\frac{1}{\tan\theta_{l_i}} - \frac{1}{\tan\theta_{r_i}}\right)$$

$$b_i = \frac{1}{2\cos(\Delta\phi_y)}\left(\frac{1}{\tan\theta_{l_i}} + \frac{1}{\tan\theta_{r_i}}\right)$$

To deal with this, we discarded all orientations within 10° of horizontal as being too unreliable.

Using the above equations, (a_i, b_i) were plotted for each trial. An example of one such plot is shown in Figure 9. Although there is spread of these values around the theoretically expected line, it is quite clear that these data can be fit with a line to recover the parameters of surface orientation, ϕ_x, ϕ_y, for a known (or assumed) value of $\Delta\phi_y$.

Some aspects of this plot can be understood by making use of a small orientation disparity approximation which is reasonable for small values of $\Delta\phi_y$. With this approximation b_i becomes $1/\tan\theta$ where θ is the mean local image orientation, $\frac{1}{2}(\theta_l+\theta_r)$. For a given uncertainty or error in orientation estimates of $\Delta\theta$, the resulting range of b_i, when θ is small, will be roughly $1/(\theta-\Delta\theta)\ldots 1/(\theta+\Delta\theta)$. This is quite significant and should clearly be taken into account when finding the line of best fit. In Figure 9, it may be noted that the points corresponding to large b_i have a greater spread.

Using the same approximation ($b_i \approx 1/\tan\theta$), the b_i axis could easily be relabelled with the corresponding mean orientation of texture elements. A $b_i = 5$ corresponds roughly to 11°, $b_i = 2$

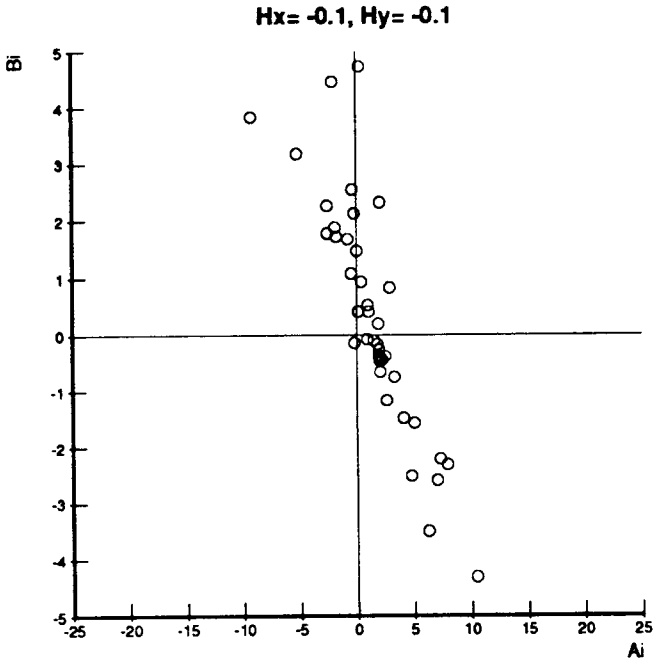


Figure 9: Typical plot of (a_i, b_i) values for one particular surface orientation. The points lie quite close to the theoretically expected line. Calculating the line of best fit provides a good estimate of the 3-D surface orientation.

to 27°, $b_i = 1$ to 45°, and $b_i = 0$ to 90° or vertical. Thus a uniform distribution of texture element orientations in the test images results in a clustering of (a_i, b_i) points towards the origin.

This same procedure was followed for all 25 test surface orientations. The resulting plots are shown in Figure 10. Plots in this grid are arranged in a manner similar to that in Figure 8, with the center plot corresponding to a fronto-parallel surface. In this case, the points clearly lie along a vertical line through the origin. The recovered parameters are $\phi_x = 0$ because the distance from the origin is zero, and $\phi_y = 0$ because the line is not rotated away from vertical. Moving left-to-right in this figure, the test surfaces were rotated around the vertical axis. The line along which the points lie clearly rotates as ϕ_y is varied. Moving top-to-bottom in this figure, the test surfaces were rotated around the horizontal axis. The signed distance from the origin clearly varies as would be expected from the change in ϕ_x (see Fig. 7).

It is important to note that this method does not strictly require the presence of explicitly identified "line" or "edge" elements. This makes the method generally applicable.

5.3 Further Work

At this point, we take these experimental results as a reassuring validation of our proposed least squares solution to recovering 3-D surface orientation from orientation disparity. We are currently performing similar tests using both orientation and spatial frequency disparities, as well as testing this scheme to recover the local surface normal on curved surfaces.

To compute ϕ_x and ϕ_y reliably, we are currently exploring line-fitting procedures (e.g., Lawson, 1974). This problem is best addressed as a total least squares problem which would treat the a_i, b_i

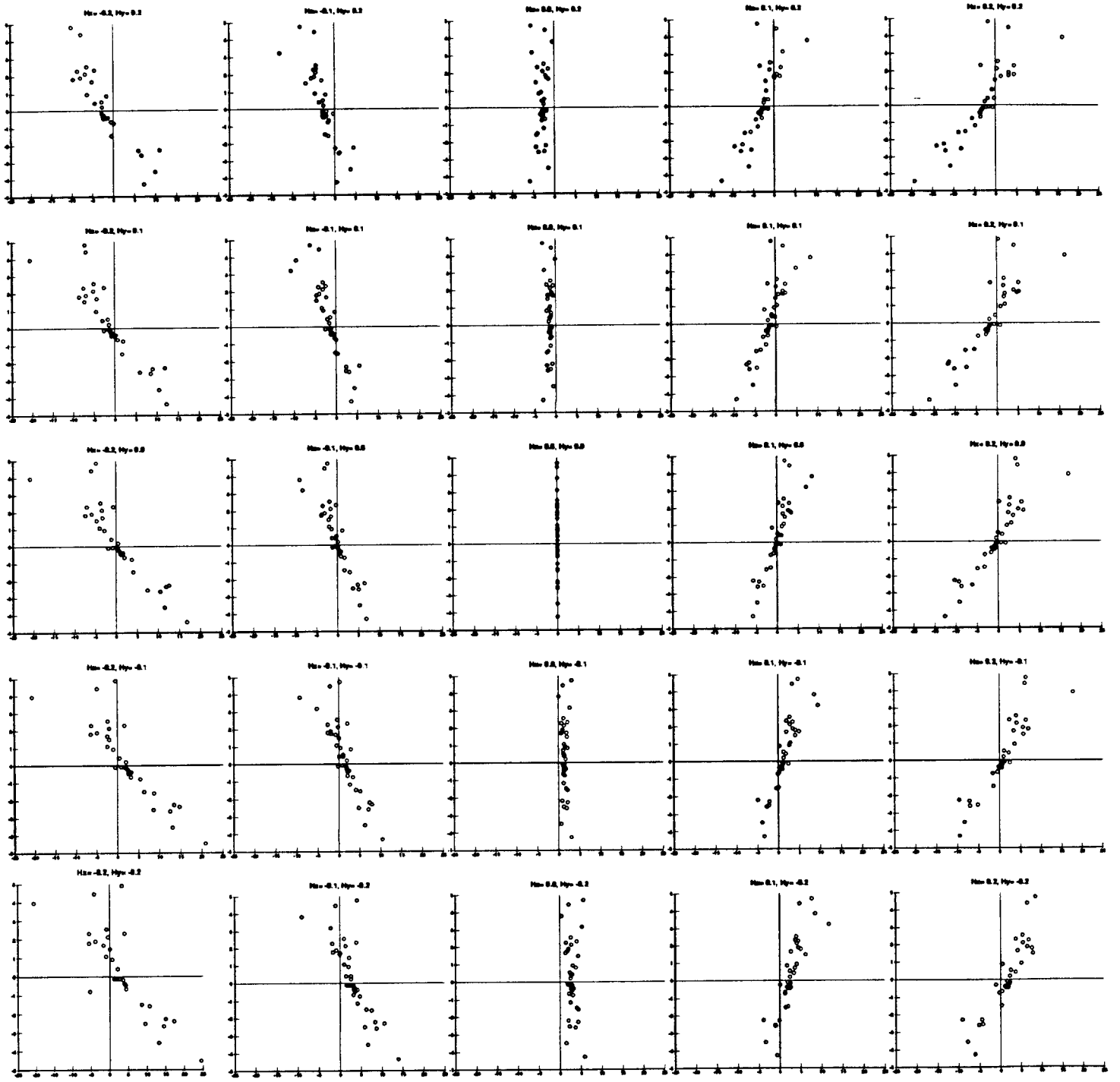


Figure 10: Plots of (a_i, b_i) for 25 test surface orientations The data clearly fall close to a line at a distance $\tan \phi_x$ from the origin, and rotated ϕ_y away from vertical.

coordinates symmetrically. Ideas from robust statistics such as least median of squares may also be explored in order to cope with occasional outliers.

Because this method is not computationally demanding, it can clearly be efficiently applied throughout an image. Where there is sufficient surface detail, the local surface normal can be computed and displayed, for example, in a "needle" diagram. We are currently investigating reliable ways for determining when it is *not* possible to make any useful assertion about the surface normal. In areas of little image intensity variation, or at boundaries between textured surfaces, the plots of (a_i, b_i) as in Figure 9 will not be well fit by a single line.

6 Conclusion

In this paper, we have developed a computational approach that exploits orientation and spatial frequency disparities to recover 3-D surface normals. The principal mathematical contribution is the derivation of constraint equations relating orientation and spatial frequency disparities to the local surface normal. These were used to derive necessary and sufficient conditions for recovering surface normals: (1) Two measurements of orientation disparity, or (2) One measurement of orientation disparity and associated spatial frequency disparity. These conditions are readily met in local regions of real images, for example in texture patches and in the neighborhood of brightness edges and lines that are curved or form corners and junctions. We have shown analytically how these constraint equations can be expressed in a form such that the determination of 3-D surface normal is equivalent to finding the best fitting straight line through a set of points. Computational experiments confirmed these results, suggesting that simple least squares fitting could be used to directly compute the local surface normals when more than the minimum number of orientation and spatial frequency disparity observations are available.

References

Arnold RD, Binford TO (1980) Geometric constraints on stereo vision. Proc. SPIE 238:281–292

Blakemore C (1970) A new kind of stereoscopic vision. Vision Res. 10:1181–1200

Grimson WEL (1981) From images to surfaces. M.I.T Press, Cambridge, Mass

von der Heydt R, Hänny P, Dursteller MR (1981) The role of orientation disparity in stereoscopic perception and the development of binocular correspondence. in Advances in Physiological Science: 16:461–470 Graystan E, Molnar P (eds) Oxford:Pergammon

Jones DG, Malik J (1990) Computational stereopsis — beyond zero-crossings. Invest Ophth and Visual Science (suppl) 31(3):529

Jones DG, Malik J (1991) Using orientation and spatial frequency disparities to recover 3–D surface shape — a computational model. Invest Ophth and Visual Science (suppl) 32(4):710

Jones DG (1991) Computational models of binocular vision. PhD Thesis, Stanford University

Jones DG Malik J (1991) A computational framework for determining stereo correspondence from a set of linear spatial filters. Technical Report UCB–CSD 91–655

Jones DG Malik J (1991) Determining three-dimensional shape from orientation and spatial frequency disparities II — using corresponding image patches. Technical Report UCB–CSD 91–657

Julesz B (1960) Binocular depth perception of computer generated patterns. Bell Syst. Tech. J. 39:1125–1162

Kass M (1983) Computing visual correspondence. DARPA Image Understanding Workshop 54–60

Kass M (1988) Linear image features in stereopsis. Int. J. Computer Vision 357–368

Koenderink JJ, van Doorn AJ (1976) Geometry of binocular vision and a model for stereopsis. Biol. Cybern. 21:29–35

Lawson CL (1974) Solving least squares problems. Prentice Hall

Mayer MJ, Kim CBY (1986) Smooth frequency discrimination functions for foveal, high-contrast, mid spatial frequencies. J Opt Soc Am A 3(11):1957–1969

Mori K, Kododi M, Asada H (1973) An iterative prediction and correction method for automatic stereo comparison. Computer Graphics and Image Processing 2:393–401

Quam LH (1984) Hierarchical warp stereo. Proc Image Understanding Workshop.

Rogers BJ, Cagenello RB (1989) Orientation and curvature disparities in the perception of 3-D surfaces. Invest Ophth and Vis Sci (suppl) 30:262

Tyler CW, Sutter EE (1979) Depth from spatial frequency difference: an old kind of stereopsis? Vision Res. 19:859-865

Watt RJ (1984) Towards a general theory of the visual acuities for shape and spatial arrangement. Vis Res 24(10):1377–1386

Wildes RP (1991) Direct recovery of three-dimensional scene geometry from binocular stereo disparity. IEEE Trans PAMI 3(8):761–774

Witkin AP, Terzopoulos D, Kass M (1987) Signal matching through scale space. Int J Computer Vision vol 2