Determining Three-Dimensional Shape from Orientation and Spatial Frequency Disparities II — Using Corresponding Image Patches

David G. Jones[†] **Jitendra Malik**[‡]

Technical Report UCB–CSD 91–657

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. Indeed, 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 machine vision algorithms, orientation and spatial frequency disparities are a source of error in finding corresponding points in left and right views, because one seeks to find features (or areas) which are similar in the two views when, in fact, they are systematically different. In other words, it is common to treat as noise what is useful signal.

We have been developing a new stereo algorithm based on the outputs of linear spatial filters at a range of orientations and scales. We present a method in this framework for making use of orientation and spatial frequency disparities to directly recover local surface slant. An implementation of this method has been tested on curved surfaces and quantitative experiments show that accurate surface orientation can be recovered efficiently. This method does not require the explicit identification of oriented line elements and also provides an explanation of the intriguing perception of surface slant in the presence of orientation or spatial frequency disparities, but in the absence of systematic positional correspondence.

[†]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 in part by a NSF 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.

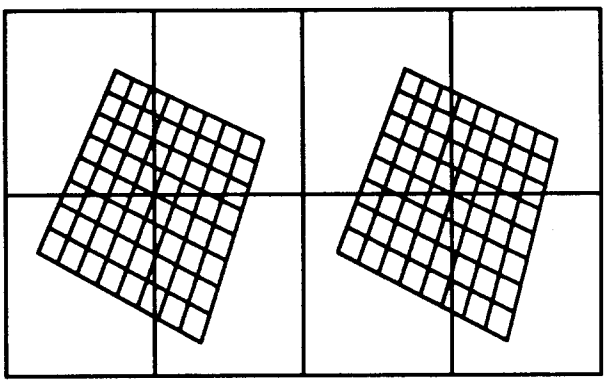

Figure 1: Stereo pair of a planar surface tilted in depth. A careful comparison of the two views reveals a slightly different orientation and spacing for corresponding grid lines drawn on the surface.

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 describes a novel computational method for directly recovering surface orientation by exploiting these orientation and spatial disparity cues. Our work is in the framework of a filtering model for computational stereopsis (Jones and Malik, 1990; Jones and Malik, 1991; Jones, 1991) where the outputs of a set of linear filters at a point are used for matching. The key idea is to model the transformation from one image to the other locally as an affine transformation with two significant parameters H_x, H_y which are the two components of the gradient of horizontal disparity. These parameters are recovered from the deformation in the outputs of the filters between corresponding points. This approach bears a clear relationship to previous work by Koenderink and Van Doorn (1976) which is discussed in Section 4.

For the special case of orientation disparity, Wildes (1991) has an alternative approach based on determining surface orientation 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, similar to that in least squares fitting, 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. Section 4 has a reformulation in terms of a locally affine transformation. Our model is specified in Section 5. Sections 6 and 7 have experimental results.

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. Using Monte Carlo methods, the *expected* magnitude and range of orientation and spatial frequency disparities have been shown to be large enough to warrant interest (Arnold and Binford, 1980; Jones, 1991). 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.

3.1 Coordinates and Parameters

In order to discuss the geometry involved 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.

3.2 Projection onto the Image Planes

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

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

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

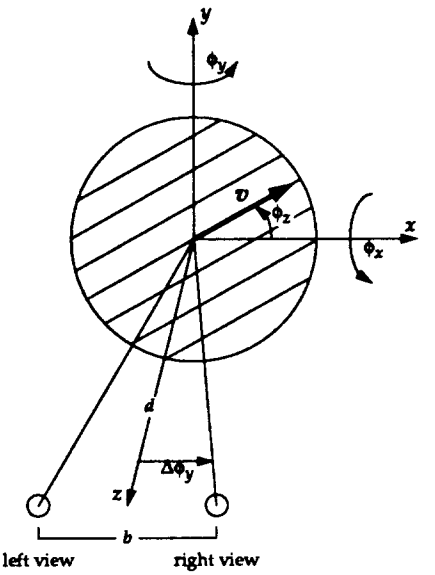

Figure 2: Parameters for specifying the three-dimensional orientation of a small planar surface patch. A planar surface (*disc*) is viewed at a distance d, from two vantage points separated by a distance b. A three-dimensional vector, $\boldsymbol{v}$, is used as a reference in the direction of a generic 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.

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

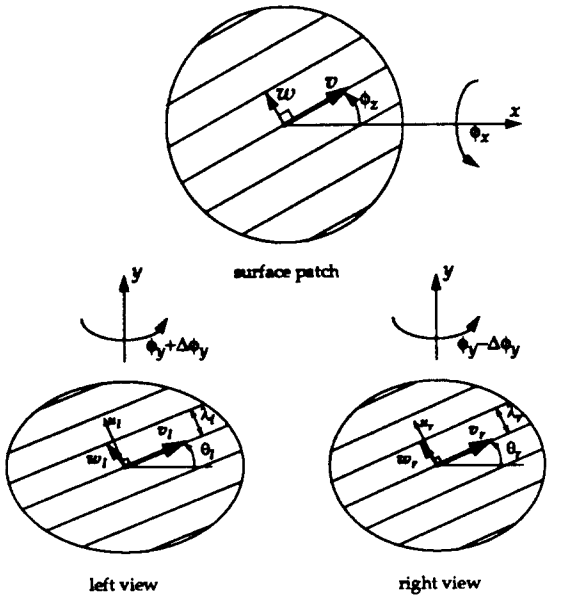

Figure 3: The slightly different appearance of a tilted surface from two viewpoints. A three-dimensional surface patch (*top*), ruled with parallel lines at an orientation given by ϕ_z, is seen from two viewpoints, giving two different images (*bottom*). For reference, three-dimensional vectors lie parallel ($\boldsymbol{v}$) 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 $\Delta\phi_y$.

For solving the inverse problem, it has been shown (Jones, 1991; Wildes, 1991) that given the orientations θ_l, θ_r and θ'_l, θ'_r of a pair of corresponding line elements, the three-dimensional surface normal can be recovered. If more orientations are available, a least squares algorithm can be used to determine the best fitting surface normal. In Section 5 we present an alternative solution which does not depend on the identification of corresponding line elements, but simply on the output of a set of linear spatial filters.

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 $\sin\theta_l$ and $\sin\theta_r$ 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. Sometimes, 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. It can be shown (Jones, 1991) that a single observation of orientation disparity and spatial frequency disparity is sufficient. 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. Where more information is available, it can be exploited using a least squares algorithm (Jones, 1991). To develop a solution in a filter-based framework, the next section first re-casts the information present in orientation and spatial frequency disparities in terms of the disparity gradient.

4 Formulation using Gradient of Horizontal Disparity

Consider a region of a surface that is visible from two viewpoints. Let $P = (x, y)$ be the coordinates of a point in this region in one image, and $P' = (x', y')$ be its corresponding point in the other image. If this surface is fronto-parallel, then P and P' differ only by horizontal and vertical offsets H, V throughout this region. In this case the image patch is one view is merely a translated version of its corresponding patch in the other view. Otherwise, if the surface is tilted or curved in depth then the corresponding image patches will not only be translated, but will also be distorted. For this discussion, it will be assumed that this distortion is well-approximated by an affine transformation.

$$P' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 + H_x & H_y \\ V_x & 1 + V_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} H \\ V \end{bmatrix}$$

H and V are the horizontal and vertical offsets or disparities due to the surface's relative depth. H_x, H_y, V_x, V_y specify the linear approximation to the distortion. When the surface is fronto-parallel, these are all zero. For planar surfaces under orthogonal projection, the transformation between corresponding image patches is correctly described by this affine transformation. For curved surfaces under perspective projection, this affine transformation provides the best linear approximation. It will turn out that the image patch over which this needs to be a good approximation is the spatial extent of the filters used.

The vertical disparity V is relatively small under most circumstances. The vertical components of the image distortion are even smaller in practice. For this reason, it will be assumed that $V_x, V_y = 0$. This leaves H_x which corresponds to a horizontal *compression* or *expansion*, and H_y which corresponds to a vertical *skew*. In both these cases, texture elements oriented near vertical are most affected. It should also be noted that the use of H_x, H_y differs from the familiar Burt-Julesz (1970) definition of disparity gradient, which is with respect to a cyclopean coordinate system.

Setting aside positional correspondence for the moment, since it has to do with relative distance to the surface and not its orientation, this leaves the following linear transformation.

$$P' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1+H_x & H_y \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

If we are interested in how a surface, or how the tangent plane to the surface, is tilted in depth, then the critical parameters are H_x and H_y. If they could be measured, then the surface orientation could be estimated, up to some factor related to the angular separation of the eyes.

What is left then is to relate these parameters in image coordinates to the previous parameters in object-centered coordinates. For a planar surface, rotated away from fronto-parallel by ϕ_x around the x-axis, and ϕ_y around the y-axis, we already have an equation that relates a point on the surface to its projection on the image. Specifically, for a point at the end of a vector with an orientation specified by ϕ_z, its orthogonal projection onto the image is given by the first equation in Section 3.2.

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

Recall that to allow for two viewpoints, ϕ_y is replaced for $\phi_y + \Delta\phi_y$ for the left view and $\phi_y - \Delta\phi_y$ for the right view. Note that since the y image coordinate does not depend on ϕ_y, it is clear that $y_l = y_r$.

$$x_r = (1+H_x) \cdot x_l + H_y \cdot y_l$$

Only two points will be needed to solve for H_x and H_y in terms of these 3-D rotations. It is convenient to choose $\phi_z = 0°$ and $\phi_z = 90°$ to get the following pair of equations.

$$\begin{aligned} \cos(\phi_y - \Delta\phi_y) &= (1+H_x) \cdot \cos(\phi_y + \Delta\phi_y) + H_y \cdot 0 \\ \sin\phi_x \sin(\phi_y - \Delta\phi_y) &= (1+H_x) \cdot \sin\phi_x \sin(\phi_y + \Delta\phi_y) + H_y \cdot \cos\phi_x \end{aligned}$$

Solving for H_x and H_y gives

$$\begin{aligned} H_x &= \frac{\cos(\phi_y - \Delta\phi_y)}{\cos(\phi_y + \Delta\phi_y)} - 1 \\ H_y &= \frac{\tan\phi_x \sin(2\Delta\phi_y)}{\cos(\phi_y + \Delta\phi_y)} \end{aligned}$$

These are the parameters for moving *from* the left view *to* the right view. To go in the other direction requires the inverse transformation. This can be computed either by changing the sign of $\Delta\phi_y$ in the above equations to to interchange the roles of the two viewpoints, or equivalently, the inverse of the transformation matrix can be computed directly.

$$\begin{aligned} H'_x &= \frac{-H_x}{1+H_x} \\ H'_y &= \frac{-H_y}{1+H_x} \end{aligned}$$

Horizontal compression or expansion, H_x, of the image in corresponding regions depends only on how far from fronto-parallel the surface is rotated around the vertical axis, ϕ_y — a rotation by ϕ_x

around the horizontal axis has no effect. Vertical skew, H_y, is only present when the surface is tilted away from fronto-parallel by a rotation around the horizontal axis, ϕ_x, but the amount of skew changes if the surface is also rotated around the vertical axis by ϕ_y.

In both cases, compression and skew depend on the angular separation $2\Delta\phi_y$ of the viewpoints and are reduced as this angle decreases. Recall that this angle is the angle subtended by the viewpoints, relative to a point on the surface, so more distant surfaces lead to a smaller angle. This means that more distant surfaces give rise to lesser amounts of compression and skew, making it more difficult to judge their inclination.

As an aside, it may be noted that humans are more sensitive to detecting vertical gradients of horizontal disparity, H_y, than horizontal gradients, H_x (Mitchison and McKee, 1990). However, we have not made any use of this in our model. It should also be noted that we motivated our particular decomposition of the linear part of the affine transformation by reasoning about the relative magnitudes of the various terms under typical binocular viewing conditions. If one cannot exploit this, as in the otherwise very similar case of optical flow, an alternative decomposition which maximizes the number of "coordinate-free" terms may be appropriate. In this case, the following formulation from (Koenderink and van Doorn, 1976) could be used:

$$\begin{bmatrix} 1+H_x & H_y \\ V_x & 1+V_y \end{bmatrix} = r\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + s\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + D$$

With this decomposition, any linear image transformation is seen to be composed of the elementary and orthogonal transformations: an anti-symmetric part, with scale factor r, a uniform dilation or contraction by s, and an area preserving deformation D. This deformation can be written as the following product

$$D = d\begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$

With D in this form, it is apparent that it has only two parameters. The angle ϕ specifies a direction and d specifies an expansion in that direction, and a corresponding contraction in the orthogonal direction. This decomposition just trades the parameters H_x, H_y, V_x, V_y for new parameters r, s, d, ϕ. It has been shown analytically that recovering the deformation component, d, ϕ, is sufficient to recover the image plane gradient of inverse radial distance (Koenderink and van Doorn, 1976). A motivation for this decomposition was to design a method insensitive to either a rotation or uniform expansion of one image because of evidence that human stereopsis tolerates up to a 6° rotation or a 15% expansion (Julesz, 1971). For small rotations, most of the change to the transformation matrix occurs in the r parameter, and for expansions, s is clearly the only parameter that changes, leaving the deformation d largely unaffected. These results on human stereopsis refer to rotations or expansions applied the entire image, however, so it is not clear whether ignoring such image distortions in localized image regions is wise.

An outline of a method for using a number of oriented line element detectors to recover the deformation component, and from that surface orientation has been proposed (Koenderink and van Doorn, 1976), but not having been implemented, the robustness of this particular approach remains unclear.

The assumption that the vertical disparity gradient is zero and that the horizontal disparity gradient provides the cue for estimating surface slant has received some support from recent psychophysical

results as the better decomposition for modelling human stereopsis. Recall that under normal stereo viewing conditions, and especially for central vision, the gradient of vertical disparity is very near zero. When an artificial stimulus with V_x non-zero is presented to a human subject, the result is an involuntary cyclo-torsion of the eyes that compensates and make the vertical disparity gradient once again zero (Rogers and Howard, 1991). For more complicated stimuli, where no single rotation of the eyes could eliminate the the vertical disparity gradients in different parts of the visual field, the non-zero V_x, V_y does not result in the perception of a slanted surface. This conflicts with the model of Koenderink and van Doorn (1976), but is in agreement with model presented here, where V_x and V_y are ignored in the recovery of surface orientation.

From a purely machine vision point of view, the advantages of neglecting the inherently small V_x, V_y are two fold: reduced computational complexity and, more importantly, improved accuracy in the detection of the remaining two parameters H_x, H_y. From numerical analysis considerations the simultaneous estimation of parameters which are widely different in magnitude is a delicate task.

4.1 The Magnitude of the Disparity Gradient

Before proceeding to make use of these image distortions, it is important to know how large they will be under different viewing conditions. Since the disparity gradient is dependent on the viewing angle, $2\Delta\phi_y$, the probability distribution of the disparity gradient is considered for two cases: typical human viewing conditions at a distance of one metre ($\Delta\phi_y = 2.0°$), and conditions more typical for aerial photography ($\Delta\phi_y = 19.3°$).

The horizontal image expansion or compression that arises when viewing a planar surface rotated ϕ_y around the vertical axis is shown in Figure 4A. Values of H_x are plotted for all angles at which the front surface is visible from both viewpoints. Recall that $1 + H_x$ is the expansion of the *right* image relative to the *left* image. Negative H_x values mean the right image is compressed relative to the left image, because the surface is slanting away to the left. Positive H_x values mean the right image is expanded relative to the left image, because the surface is slanting away to the right. Note that H_x as a function of ϕ_y is not symmetric: $1 + H_x$ has a minimum value of 0, ($H_x = -1$), and a maximum value of ∞.

A useful case to consider is a surface rotated 45° around the vertical axis ($\phi_y = \pm 45°$). In this case, $H_x \approx \pm 0.07$ under reasonable human viewing conditions. A 7% difference is well within the range to be used by the human visual system. The shallow slope of this function near fronto-parallel, however, suggests that rather subtle differences in compression/expansion would need to be discerned to discriminate small differences in tilt. For the same surface slant, but under conditions more typical for aerial photography, $H_x \approx -0.5$ or $+ 1.0$. These values are quite large, suggesting this information could easily be used in an application with a large stereo baseline. Furthermore, these large image distortions would clearly pose difficulties for approaches that ignore them by assuming the corresponding image patches are translated copies of one another.

The vertical image skew that arises when viewing a planar surface rotated ϕ_x around the horizontal axis is shown in Figure 4B. Positive H_y values mean that the right image patch is skewed counter-clockwise relative to the left image. Notice that H_y as a function of ϕ_x is symmetric when $H_x = 0$. In Figure 4C, where $H_x = H_y$, this is no longer the case. This figure also illustrates that a rotation about the horizontal axis can increase the image skew.

When the disparity gradient (H_x, H_y) becomes too large, it may be impossible to measure accurately

because corresponding image patches are too dissimilar. An upper limit on the size of the allowable disparity gradient has been incorporated in several models of stereo vision (e.g., Pollard et al., 1985). Note that this sets limits on the range of surface tilts that can be measured, and these limits are different for different viewing conditions. For example, if a disparity gradient limit of 1.0 is chosen, the maximum tolerable surface slant is still quite large ($\phi_x, \phi_y \approx \pm 85°$) under typical human stereo viewing conditions. For a larger baseline, however, the maximum tolerable surface slant significantly reduced ($\phi_x \approx \pm 56° \phi_y \approx \pm 45°$). This would be an important consideration for aerial photography of a mountainous region.

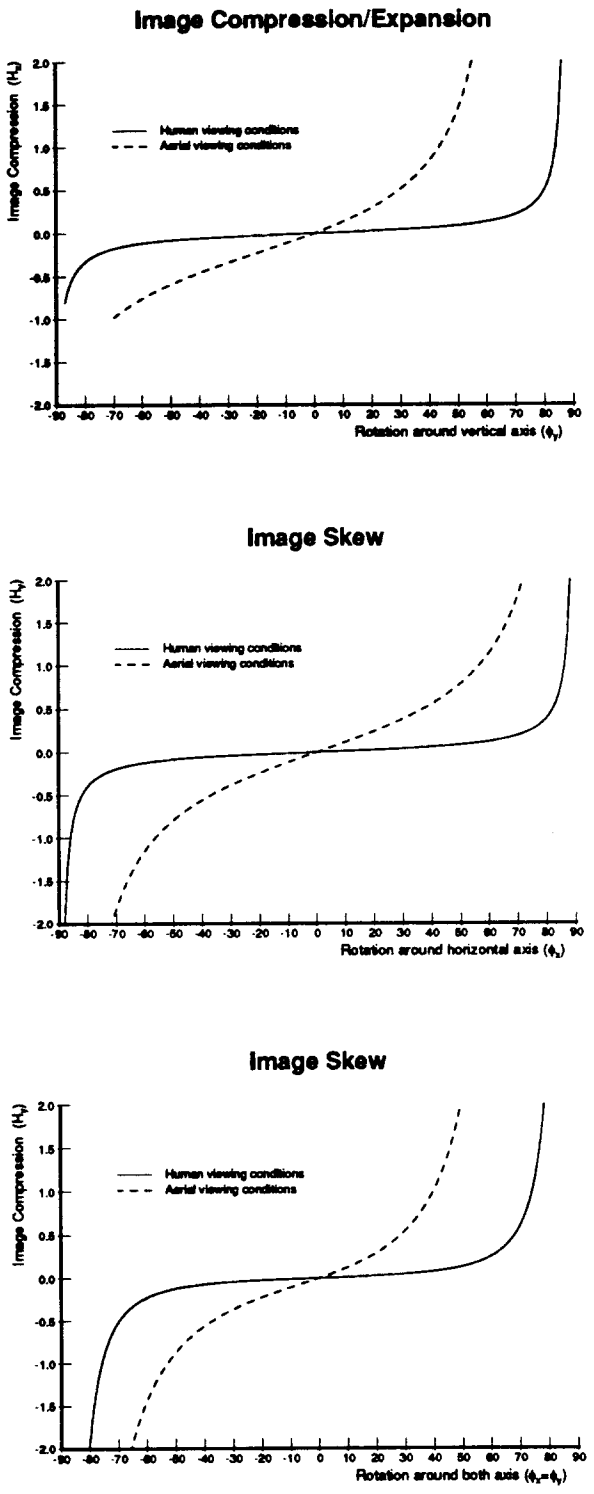

Figure 4: Image distortions present when viewing a surface tilted in depth. (A) The horizontal image compression, H_x, resulting when the surface being viewed has been rotated, ϕ_y, around the vertical axis. (B) The vertical image skew, H_y, resulting when the surface being viewed has been rotated, ϕ_x, around the horizontal axis. (C) The vertical image skew resulting when the surface being viewed has been rotated around both the horizontal and vertical axes by the same amount ($\phi_x = \phi_y$).

4.2 The Expected Range of the Disparity Gradient

Monte Carlo simulation was used to determine the probability distribution for the disparity gradient (H_x, H_y) with the assumption that all surface texture orientations are equally probable and all surface orientations visible from the two viewpoints are also equally probable. One million random trials were used.

The probability distribution for H_x, which corresponds to horizontal image compression or expansion, is shown in Figure 4A. For both human and aerial photography viewing conditions, the distributions are skewed, though it is more apparent in the latter case. The median value of H_x, however, should still be exactly zero, since a surface slant to one side or the other are equally probable. The simulation bears this expectation out: the median value for H_x is 0.0002 for human viewing conditions ($\Delta\phi_y = -2.0°$) and -0.0011 for aerial viewing conditions ($\Delta\phi_y = 19.3°$).

For various percentage intervals around the median, the range of H_x is given in the table below. For example, the last entry means that 95% of the time, H_x will lie within the range -0.475, $+0.907$ for human viewing conditions and within the range -0.909, $+9.942$ for aerial viewing conditions.

Horizontal Image Expansion or Compression (H_x)

percent	human vision $\Delta\phi_y = 2.0°$	aerial photography $\Delta\phi_y = 19.3°$
25	-0.028, $+0.029$	-0.201, $+0.250$
50	-0.065, $+0.070$	-0.397, $+0.659$
75	-0.146, $+0.171$	-0.635, $+1.737$
90	-0.310, $+0.450$	-0.828, $+4.804$
95	-0.475, $+0.907$	-0.909, $+9.942$

The probability distribution for H_y, which corresponds to vertical image skew, is shown in Figure 4B. These symmetric distributions have experimentally determined medians of 0.0002 and 0.0007 for human and aerial viewing conditions respectively. The expected range of H_y for various percentage intervals centered at the median are given in the table below.

Vertical Image Skew (H_y)

percent	human vision $\Delta\phi_y = 2.0°$	aerial photography $\Delta\phi_y = 19.3°$
25	± 0.028	± 0.227
50	± 0.068	± 0.531
75	± 0.158	± 1.189
90	± 0.380	± 2.819
95	± 0.691	± 5.407

5 Surface Shape from Differences in Spatial Filter Outputs

Consider a three-dimensional surface patch, depicted at the top of Figure 6. Because of the two slightly different viewpoints, the left and right views will be slightly different. If it can be assumed

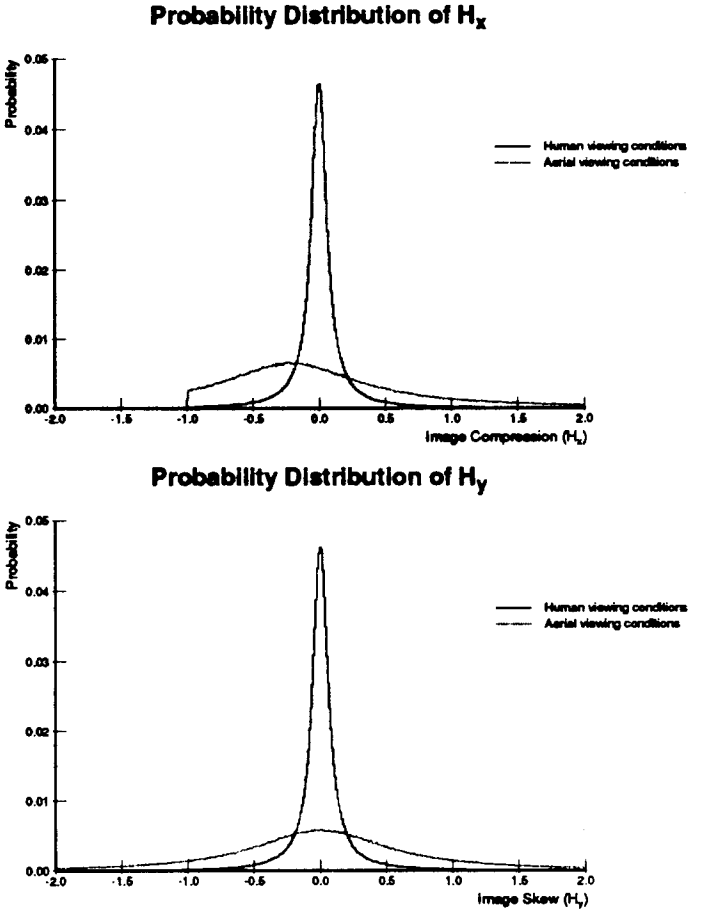

Figure 5: Probability distribution of the disparity gradient (H_x, H_y). Assuming a uniform random distribution of surface orientations, the probability distributions for H_x (A) and H_y (B) are plotted, following a Monte Carlo simulation with 1,000,000 trials.

the surface is locally planar, the transformation between these images is described by an affine transformation. Setting aside for the moment the translational component H, V, what is left is a linear transformation with parameters H_x, H_y. If the pixels of each image are listed in a column vector, then this transformation which is in terms of image coordinates, can equally well be expressed as a linear transformation from one image vector to another.

$$\begin{bmatrix} \\ I_L \\ \\ \end{bmatrix} = \begin{bmatrix} \\ T_{H_x,H_y} \\ \\ \end{bmatrix} \begin{bmatrix} \\ I_R \\ \\ \end{bmatrix}$$

In the continuous case, this would be exact, but because the images are sampled at discrete locations (pixels), this transformation is an approximation. In some instances, pixels from one image are mapped to locations in the other image which have fractional coordinates and the results must be interpolated. For all but very high frequency patterns, this interpolation is quite good. This transformation is the image warping step that is used in some other approaches mentioned earlier that cope with image distortions.

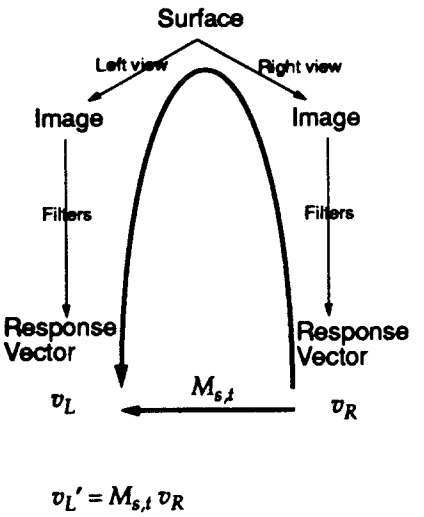

Figure 6: Comparing spatial filter outputs to recover 3-D surface orientation.

Referring to Figure 6 again, consider just the right view. This image region is filtered by a set of nearly independent linear spatial filters. Each filter f_i can be written as a column vector so the response of the filter is simply the dot product with the image vector. The entire ensemble of filters can be combined into a large matrix F and the responses of all the spatial filters can be listed in a single "filter response vector" v_R.

$$\begin{bmatrix} \\ v_R \\ \\ \end{bmatrix} = \begin{bmatrix} \\ F^T \\ \\ \end{bmatrix} \begin{bmatrix} \\ I_R \\ \\ \end{bmatrix}$$

If the surface being viewed is nearly fronto-parallel, then v_L and v_R will be very similar. If the surface is tilted in depth, then they will differ, but in a way related to surface orientation. In what follows, a method is described for determining which choice of H_x, H_y best accounts for the differences between these filter responses vectors.

Using a singular value decomposition, F^T can be decomposed into a change of basis matrix, U, a diagonal matrix of singular values or weights, Σ, and a matrix V^T the rows of which form an orthonormal basis set for the vector space spanned by the set of spatial filters. If any filters are redundant, this procedure will make it apparent because some of the singular values will be zero.

$$\begin{bmatrix} \\ F^T \\ \\ \end{bmatrix} = \begin{bmatrix} \\ U \\ \\ \end{bmatrix} \begin{bmatrix} \\ \Sigma \\ \\ \end{bmatrix} \begin{bmatrix} \\ V^T \\ \\ \end{bmatrix}$$

This decomposition can be used to determine the pseudo-inverse of the spatial filters, which could be used to reconstruct the original image patch as best as possible from the information in the filter responses.

$$\begin{bmatrix} \\ I'_R \\ \\ \end{bmatrix} = \begin{bmatrix} \\ V \\ \\ \end{bmatrix} \begin{bmatrix} \\ 1/\Sigma \\ \\ \end{bmatrix} \begin{bmatrix} \\ U^T \\ \\ \end{bmatrix} \begin{bmatrix} \\ v_R \\ \\ \end{bmatrix}$$

This procedure is numerically stable unless some of the singular values are very small, but this does not occur when the spatial filters are largely independent.

For a particular candidate surface orientation, specified by H_x, H_y, the reconstructed image patch can be transformed to give an estimate of how the surface would look from the other viewpoint.

$$\begin{bmatrix} \\ I'_L \\ \\ \end{bmatrix} = \begin{bmatrix} \\ T_{H_x,H_y} \\ \\ \end{bmatrix} \begin{bmatrix} \\ I'_R \\ \\ \end{bmatrix}$$

Lastly, filtering this transformed image patch gives a prediction of the filter outputs for the other view. As depicted by the large arc in Figure 6 it is possible to start from the filter responses in one view and, for a given choice of H_x, H_y, calculate a prediction of what the filter responses in the other view are expected to be, by this sequence of linear transformations.

$$\begin{bmatrix} \\ v'_L \\ \\ \end{bmatrix} = \underbrace{\begin{bmatrix} \\ F^T \\ \\ \end{bmatrix} \begin{bmatrix} \\ T_{H_x,H_y} \\ \\ \end{bmatrix} \begin{bmatrix} \\ V \\ \\ \end{bmatrix} \begin{bmatrix} \\ 1/\Sigma \\ \\ \end{bmatrix} \begin{bmatrix} \\ U^T \\ \\ \end{bmatrix}}_{\begin{bmatrix} \\ M_{H_x,H_y} \\ \\ \end{bmatrix}} \begin{bmatrix} \\ v_R \\ \\ \end{bmatrix}$$

Of course all of these linear transformations can be collapsed into a single one M_{H_x,H_y} which maps the filter responses from one view directly to a prediction for the filter responses in the other view. These M matrices which depend only on H_x and H_y and not the input images, can of course be pre-computed once and for all ahead of time.

There now provides a simple procedure for estimating the disparity gradient (surface orientation) directly from v_R and v_L, the output of linear spatial filters. For a variety of choices of H_x, H_y, compare $v'_L = M_{H_x,H_y} \cdot v_R$, the filter responses predicted for the left view, with v_L, the filter responses actually measured for the left view. The choice of H_x, H_y which minimizes the difference between v'_L and v_L is the best estimate of the disparity gradient. The sum of the absolute differences between corresponding filter responses serves as a efficient and robust method for computing the *difference* between these two vectors, or an error-measure for each candidate H_x, H_y.

An alternative approach for recovering affine transformation parameters from changes in filter outputs in the context of optical flow is due to Werkhoven and Koenderink (1990). Unlike the

finite displacement case considered here, they make use of a differential approximation which enables them to obtain a closed form solution of the parameters. We believe that the elegance of obtaining an analytic solution is outweighed by the numerical problems inherent in making infinitesimal approximations. In the context of stereopsis, this problem becomes particularly acute; the eyes (or cameras) are at fixed finite distances. This makes such a differential approach inappropriate.

5.1 Efficiency and Coarse Coding

The fact that the transformation matrices can be pre-computed ahead of time is crucial for this method to be practical. Still, if many choices of H_x, H_y need to be considered, that is a lot of computation just for estimating two parameters, especially considering that in the end H, V also need to be estimated. Unfortunately, these transformations are not a separable functions of H_x and H_y, so it is not possible to do two one-dimensional searches to first solve for H_x and then for H_y.

For a particular position in an image of a textured surface tilted in depth, consider the error measures for each choice of H_x, H_y. This error surface is always quite smooth. This is a natural consequence of the fact that the spatial filters used (Gaussian derivatives), and the family of transformations between image patches are both smooth functions. This means that this two dimensional parameter space can be sampled quite coarsely, with the final estimate of the parameters being interpolated without considerable loss in accuracy. This coarseness can be chosen to achieve a desired tradeoff between accuracy and efficiency, with little risk that the correct answer will "fall through the cracks" between samples.

A biologically plausible implementation of this model would be based on units coarsely tuned in positional disparity, as well as the two parameters of surface slant. The necessary types of computation amount to no more than a weighted sum of linear spatial filter outputs followed by absolute differences, and these weights are fixed.

5.2 An Example of Recovering Surface Shape

Up to this point, the discussion has been purely theoretical. An algorithm has been developed that, at each point in an image, determines the horizontal and vertical components of positional disparity H, V, as well the two components of the horizontal disparity gradient, H_x, H_y. This gives the relative distance and surface orientation for each point in the scene.

There are few potential concerns that one might have when implementing this algorithm. First, some approximations are involved in the calculation of the *predicted* filter response vector that must be compared with the observed one. The pseudo-inverse, for example, only partially reconstructs the image patch encoded by each filter response vector. The extent to which this does or does not cause a problem will naturally depend on whether the particular filter set employed has encoded enough of the relevant information. For the task of recovering surface orientation, the filters must span a large enough range of orientations and spatial frequencies.

Another question that is easiest answered by testing an implementation of this algorithm is the degree to which deviations from the assumption that surfaces are locally planar can be tolerated. For this reason, this algorithm has been tested on a stereo pair of a textured sphere (Fig. 7).

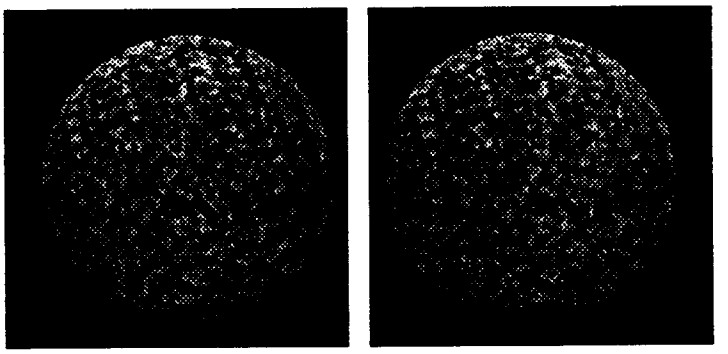

Figure 7: Stereo pair of a textured sphere.

A table of matrices, M_{H_x,H_y} was computed ahead of time for 11 discrete H_x, H_y values. At each point in the image, the four parameters H, V, H_x, H_y that minimized the difference between the predicted and observed filter response vectors was recorded. The results are shown in Fig. 8. For simplicity, none of the iterative refinement techniques discussed earlier were used. Also, no interpolation was used to improve upon the somewhat coarse H_x, H_y values. In the next section, the benefits of such an interpolation is clearly demonstrated.

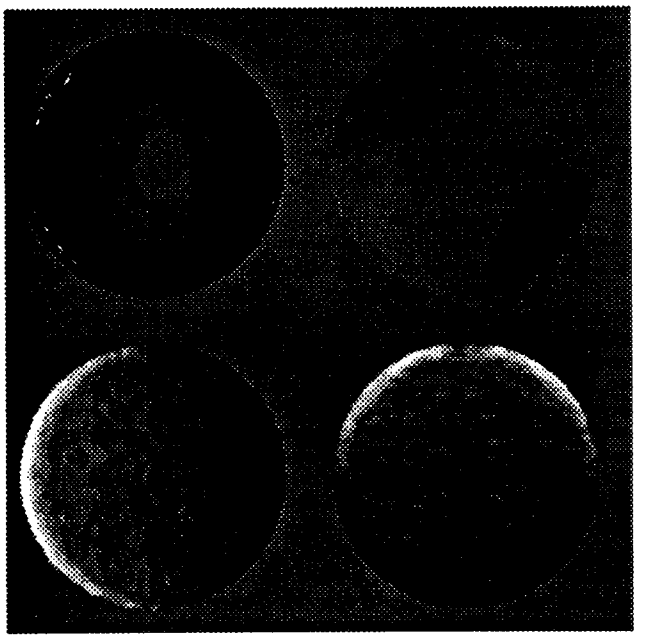

Figure 8: Recovered disparity maps for textured sphere. The two components of positional disparity are shown at the top: horizontal disparity, H, (*top left*), and vertical disparity V, (*top right*). The two components of the disparity gradient are shown at the bottom: H_x (*bottom left*), H_y (*bottom right*). Zero values are shown as middle grey. Positive values are lighter and negative values are darker.

The horizontal and vertical disparity estimates are quite accurate, except on a very thin crescent which was only visible in one view. It is also easy to see that the disparity gradient estimates are also quite good. The center of the sphere is fronto-parallel ($H_x = 0$, $H_y = 0$), and the H_x values are increasing negative on one side of the sphere and increasingly positive on the other. Similarly,

the recovered H_y values accurately reflect the vertical component of the surface orientation.

6 The Accuracy of Recovered Surface Orientations

Though trying this algorithm on example stereo pairs and examining the results by eye can help demonstrate the basic successful performance of this scheme for recovering local surface orientation, it is instructive to have a more quantitative assessment. To this end, experiments were performed to determine how accurately this approach can estimate the the parameters related to surface orientation.

6.1 Experimental Method

The computer implementation of this approach was tested quantitatively using test stereo pairs such as those shown in Figure 9. Each of these nine stereo-pairs depict 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 square marked on the planar surface makes the horizontal compression/expansion, when $H_x \neq 0$, and vertical skew, when $H_y \neq 0$, quite apparent.

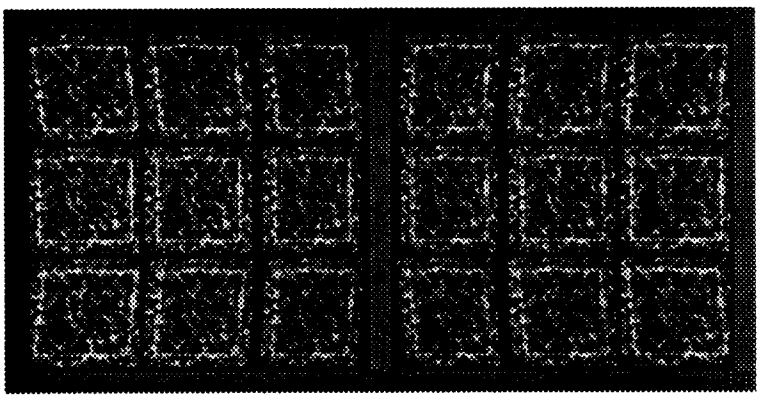

Figure 9: Stereo pair of test surfaces tilted in depth.

In total, 49 test surface orientations were used, with H_x, H_y taking on values: $0.0, \pm 0.1, \pm 0.2, \pm 0.4$. For each test surface orientation, 50 stereo-pairs of a textured, planar surface were created. On each trial, the parameters of surface orientation, H_x, H_y were computed using the method described in this paper. This resulting estimate of surface orientation was based solely upon filter outputs from the central point in the image. If surface orientation was estimated by averaging over several positions in the image, the resulting estimate would more accurate. The procedure followed here thus provides a conservative evaluation of this approach.

6.2 Experimental Results

The graph in Figure 10 shows the disparity gradient estimates, computed by the model, for the various test surfaces. The axes are the parameters of surface orientation, H_x, H_y, with $(0, 0)$ at the

center being fronto-parallel. Each cluster of dots shows the results of 50 trials. The central open dot is the test surface orientation and the black dots are the surface orientation recovered on each trial.

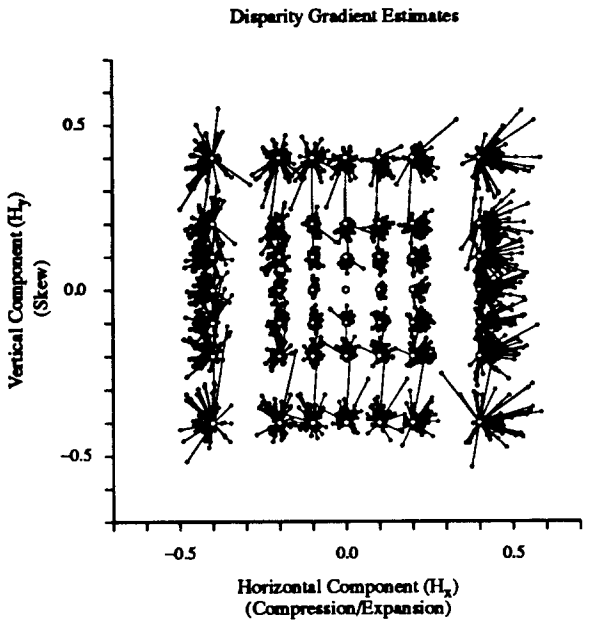

Figure 10: Disparity gradient estimates for various test surfaces.

These data are summarized in Figure 11, where each test surface orientation is again marked by an open dot, and the centroid and standard deviation of the recovered surface slant parameters are marked by a black dot and ellipse. (The standard deviation is also known as the Mahalonobis distance.) The recovered surface orientations are quite accurate, especially for small slants. For larger slants, the spread in the recovered surface orientation increases, which is agreement with human psychophysics since our ability to discriminate different surface orientations declines as the slant increases.

There are some small systematic errors. For example, for large rotations around the vertical axis, the slant was overestimated. To the extent that this is as smooth function of surface orientation, as it appears in these results, this could easily be compensated for by a simple adjustment to the recovered H_x, H_y values. It is not clear, however, whether this is an inherent feature of the model, or whether it is an artifact of this particular implementation. Estimating surface orientation from coarse estimates using the parabolic interpolation as was done here is likely to give rise to some systematic errors. By examining the general nature of the error surface in the vicinity of actual surface orientations it should be possible to improve the interpolation accuracy, as well as provide a better understanding of just how many coarse samples are needed for a desired level of accuracy.

This evaluation provides a conservative estimate of the accuracy with which this approach can recover surface orientation. The data presented here are surface orientation estimates taken at a single position in the image. Because the test surfaces are marked with random textures, the orientation and spatial frequency disparities at this single position encode surface orientation to varying degrees, and on some trials would provide only very limited cues. For example, horizontal stripes provide *no* information about rotation around the vertical axis. For large planar surfaces, or smooth surfaces in general, estimates could be substantially refined by pooling over a local

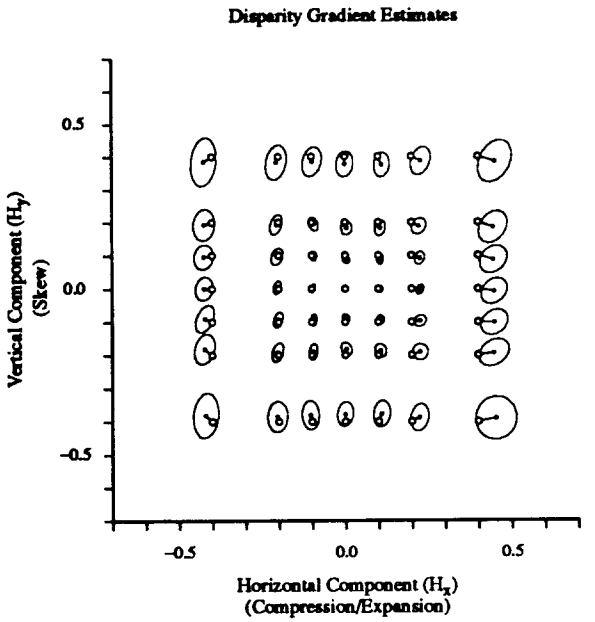

Figure 11: Mean and standard deviation of disparity gradient estimates.

neighborhood. This would, however, trade off spatial resolution for this increased accuracy.

7 Orientation and Spatial Frequency Disparities without Correspondence

The approach for recovering three-dimensional surface orientation developed here makes use of the fact that it is the *identical* textured surface patch that is seen in the two views. It is this assumption of correspondence that allows an accurate recovery of the parameters of the deformation between the two retinal images. Recall, however, that in Section 2 experiments due to von der Heydt (1981), and Tyler and Sutter (1979), were described in which orientation and spatial frequency disparities lead to the perception of a tilted surface, in the *absence* of any systematic correspondence. One interpretation of those results might suppose the existence of stereo mechanisms which make use of orientation or spatial frequency disparities *independent* of positional disparities or correspondence. Such mechanisms would seem to be necessarily quite different from the approach suggested here. On the other hand, it is not immediately apparent how the present approach would perform in the absence of correspondence.

From the previous experiment, we have an implementation that determines, given a pair of images, the best estimate of surface orientation. It will always find a best estimate — even if it is nonsense. This means that the same experiments described in Section 2 can be performed, except instead of a human subject, the subject will be our computational model.

7.1 Orientation Disparity

Figure 12 shows an example stimulus which contains an orientation disparity in the two images, but as inspection of this image pair will confirm, there is no systematic correspondence. Each image was created by filtering random, uncorrelated one-dimensional noise to have a bandwidth of 1.2 octaves. This is the same bandwidth used in the psychophysical experiments described earlier. In this example, there is a 10° orientation disparity. Since a different random seed is used for each image, there is no consistent correspondence or phase relationship between the two images.

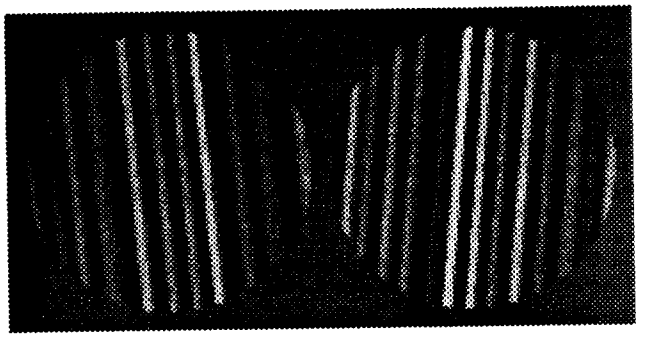

Figure 12: Orientation Disparity without Correspondence.

A sequence of 100 such pairs were created and for each, using an the same implementation of the model used in the previous experiment, the parameters of surface orientation, or the disparity gradient, H_x, H_y were estimated. Once again, this estimate was made by comparing the outputs of the linear spatial filters only at the central pixel in each image. As discussed before, this gives a rather conservative estimate of the the performance of this approach.

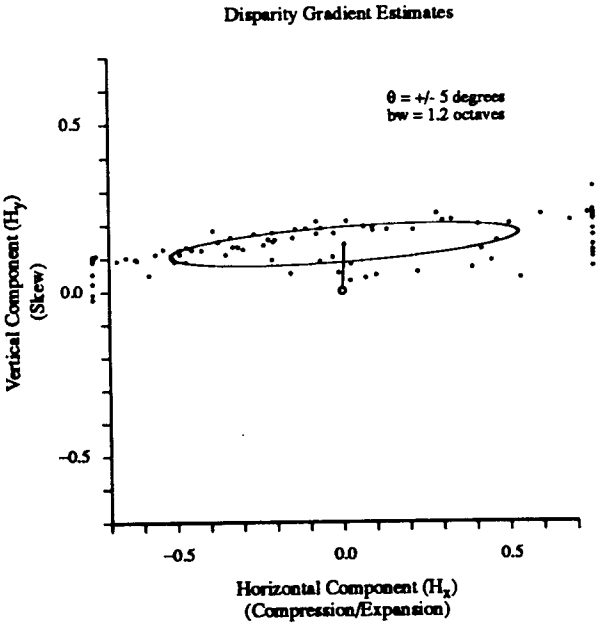

Figure 13: Disparity gradient estimates from orientation disparity alone.

These estimated parameters are shown in Figure 13. There is a fair bit of scatter in these estimates, but if the image pairs were presented rapidly, one after the other, as in the experiment, one

might expect the perceived surface slant to be near the centroid. In this case, $H_x = 0$ and H_y is positive, which corresponds to a surface rotated around the horizontal axis — in agreement with von der Heydt's results (von der Heydt et al., 1981). In fact, the centroid lies close to where it should be based on the 10° orientation disparity ($H_x = 0.0, H_y = 0.175$), despite the absence of correspondence.

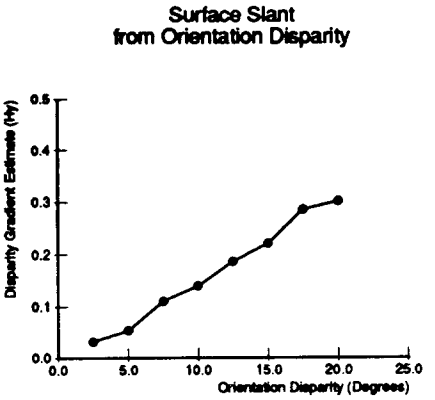

Figure 14: Disparity gradient estimates from orientation disparity alone.

The same procedure was repeated for several different orientation disparities and the results are shown in Figure 14. This graph shows the vertical component of the disparity gradient, H_y as a function of orientation disparity. In each of these horizontal component, H_x, was always near zero. For a considerable range of orientation disparities, the recovered slant increases with orientation disparity. This is also in agreement with the psychophysical results of von der Heydt.

7.2 Spatial Frequency Disparity

A similar experiment was carried out for the case of spatial frequency disparity in the absence of correspondence Figure 15 shows an example image pair with a spatial frequency disparity, but no systematic correspondence. This stereo-pair was be created in a manner very similar to that used in the preceding experiment. It is uncorrelated one-dimensional noise, filtered to have a bandwidth of 1.2 octaves and in this case, a spatial frequency disparity. In this example, they differ in frequency by a factor of 1.4.

The results are shown in Figure 16. This graph shows the horizontal component of disparity gradient, H_x, corresponding to a rotation around the vertical axis, as a function of spatial frequency ratio. Once again, over a range of spatial frequency disparities, even though there is no systematic positional correspondence, the recovered surface slant increases with increasing spatial frequency disparity.

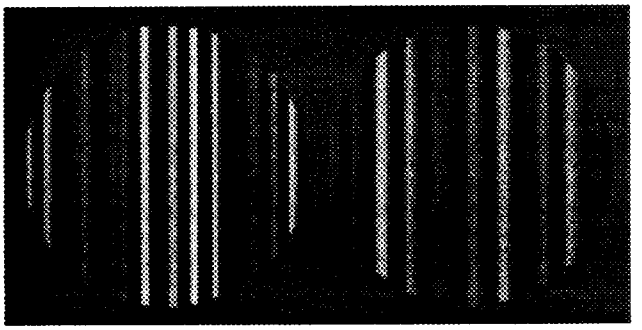

Figure 15: Spatial frequency disparity without correspondence.

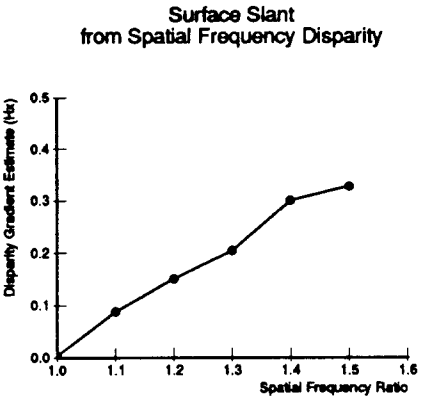

Figure 16: Disparity gradient estimates as a function of spatial frequency disparity.

8 Conclusion

In this paper, a simple stereopsis mechanism, based on using the outputs of a set of linear spatial filters at a range of orientations and scales, has been proposed for the direct recovery of local surface orientation. Tests have shown it is applicable even for curved surfaces, and that interpolation between coarsely sampled candidate surface orientations can provide quite accurate results. Estimates of surface orientation are more accurate for surfaces near fronto-parallel, and less accurate for increasing surface slants.

There is also good agreement with human performance on artificial stereo pairs in which systematic positional correspondence has been eliminated. This suggests that the psychophysical results involving the perception of slant in the absence of correspondence may be viewed, not as an oddity, but as a simple consequence of a reasonable mechanism for making use of positional, orientation, and spatial frequency disparities to perceive three-dimensional shape.

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

Burt P, Julesz B (1980) A disparity gradient limit for binocular function. Science 208:651–657

Drumheller M, Poggio T (1986) On parallel stereo. Proc IEEE Int Conf on Robotics and Automation

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

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

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

Julesz B (1971) Foundations of cyclopean perception. University of Chicago Press:Chicago

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

Mitchison GJ, McKee SP (1990) Mechanisms underlying the anisotropy of stereoscopic tilt perception. Vision Res. 30(11):1781–1791

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

Pollard SB, Mayhew JEW, Frisby JP (1985) PMF: a stereo correspondence algorithm using a disparity gradient limit. Perception 14:449–470

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

Rogers BJ, Howard IP (1991) Differences in the mechanisms used to extract 3-D slant from disparity and motion parallax cues. Invest Ophth and Vis Sci (suppl) 32:695

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

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

Werkhoven P, Koenderink JJ (1990) Extraction of motion parallax structure in the visual system 1. Biol. Cybern.

Werkhoven P, Koenderink JJ (1990) Extraction of motion parallax structure in the visual system 2. Biol Cyb

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