Determining Three-Dimensional Shape from Orientation and Spatial Frequency Disparities *

David G. Jones[1] *and Jitendra Malik*[2]

[1] McGill University, Dept. of Electrical Engineering, Montréal, PQ, Canada H3A 2A7
[2] University of California, Berkeley, Computer Science Division, Berkeley, CA USA 94720

Abstract. Binocular differences in orientation and foreshortening are systematically related to surface slant and tilt and could potentially be exploited by biological and machine vision systems. Indeed, human stereopsis may possess a mechanism that specifically makes use of these orientation and spatial frequency disparities, in addition to the usual cue of horizontal disparity. In machine vision algorithms, orientation and spatial frequency disparities are a source of error in finding stereo correspondence 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, 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.

1 Introduction

Stereopsis has traditionally been viewed as a source of depth information. In two views of a three-dimensional scene, small positional disparities between 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 $z(x, y)$. In practice, any inaccuracies present in disparity estimates will be compounded by taking derivatives.

However, there are other cues available under binocular viewing that can provide direct information about surface orientation. When a surface 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. It has been demonstrated that humans are able to exploit these cues, when present, to more

* 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.

accurately determine surface orientation (Rogers and Cagenello, 1989). In stimuli consisting of uncorrelated dynamic visual noise, filtered to contain a certain spatial frequency band, the introduction of a spatial frequency disparity or orientation disparity leads to the perception of slant, despite the absence of any systematic positional disparity cue (Tyler and Sutter, 1979; von der Heydt et al., 1981). In much the same way that random dot stereograms confirmed the existence of mechanisms 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 orientation and spatial frequency disparities in the two retinal images to aid in the perception of surface shape.

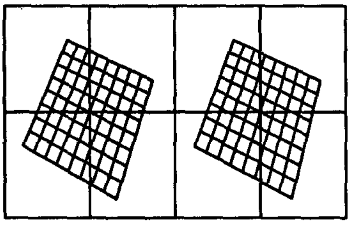

Fig. 1. Stereo pair of a planar surface tilted in depth. Careful comparison of the two views reveals slightly different orientation and spacing of corresponding grid lines.

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; Kass, 1987). A few approaches seek to cope using an iterative framework, starting from an initial assumption that disparity is locally constant, and then guessing at the parameters of the image distortion to locally transform and compensate so that image regions can again be compared under the assumption that corresponding regions are merely translated copies of one another (Mori et al., 1973; Quam, 1984; Witkin et al., 1987). The reliance of this procedure on convergence from inappropriate initial assumptions and the costly repeated "warping" of the input images make this an unsatisfactory computational approach and an unlikely mechanism for human stereopsis.

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 filter-based model for computational stereopsis (Jones, 1991; Jones and Malik, 1992) 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, the gradient of horizontal disparity. Previous work has sought to recover the deformation component instead (Koenderink and van Doorn, 1976).

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 advantage that it treats both orientation and spatial frequency disparities. Another benefit, similar to least squares fitting, it makes 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.

2 Geometry of Orientation and Spatial Frequency Disparities

Consider the appearance of a small planar surface patch, ruled with a series of evenly spaced parallel lines (Fig. 2A). The results obtained will apply when considering orientation and spatial frequencies of general texture patterns. To describe the parameters of an arbitrarily oriented plane, start with a unit vector pointing along the x-axis. A rotation ϕ_z around the z-axis describes the orientation of the surface texture. Rotations of ϕ_x around the x-axis, followed by ϕ_y around the y-axis, combine to allow any orientation of the surface itself. The three-dimensional vector $\boldsymbol{v}$ resulting from these transformations indicates the orientation of the lines ruled on the surface and can be written concisely:

$$\left[\begin{array}{c} \\ \boldsymbol{v} \\ \\ \end{array}\right] = \left[\begin{array}{c} \sin\phi_x \sin\phi_y \sin\phi_z + \cos\phi_y \cos\phi_z \\ \cos\phi_x \sin\phi_z \\ \sin\phi_x \cos\phi_y \sin\phi_z - \sin\phi_y \cos\phi_z \end{array}\right]$$

In order to consider orientation and spatial frequency disparities, this vector must be projected onto the left and right image planes. In what follows, orthographic projection will be used, since it provides a very close approximation to perspective projection, especially for the small surface patches under consideration and when line spacing is small relative to the viewing distance. The projection of $\boldsymbol{v}$ onto the left image plane is achieved by replacing ϕ_y with $\phi_y + \Delta\phi_y$ (where $\Delta\phi_y = \tan^{-1}(b/2d)$), and then discarding the z component to give the two-dimensional image vector $\boldsymbol{v}_l$. Similarly, replacing ϕ_y with $\phi_y - \Delta\phi_y$ gives $\boldsymbol{v}_r$, the projection of $\boldsymbol{v}$ on the right image plane.

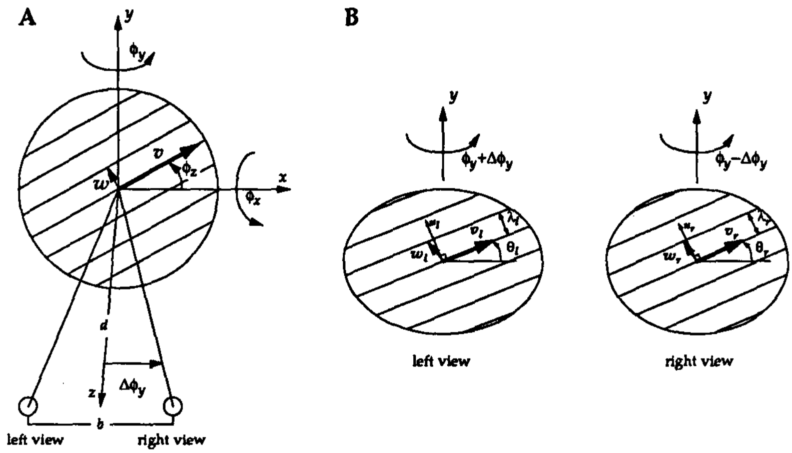

Fig. 2. Differences in two views of a tilted surface. *A*. A planar surface is viewed at a distance d, from two vantage points separated by a distance b. Three-dimensional vectors lie parallel (v) and perpendicular (w) to a generic surface texture (*parallel lines*). Arbitrary configurations are achieved by rotations ϕ_z, ϕ_x, and ϕ_y, in that order. Different viewpoints are handled by adding an additional rotation $\pm\Delta\phi_y$, where $\Delta\phi_y = \tan^{-1}(b/2d)$. *B*. Resulting two-dimensional image textures are described by orientation, θ, and spacing, λ. Orientation disparity, $\theta_r - \theta_l$, and spatial frequency disparity, λ_l/λ_r, are systematically 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. 2B). These can be easily expressed in terms of the components of the image vectors.

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

This enables us to determine the orientation disparity, $\theta_r - \theta_l$, given the pattern orientation ϕ_z, 3-D surface orientation characterized by ϕ_x, ϕ_y, and view angle $\Delta\phi_y$.

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. 2B). Since spatial frequency is measured perpendicular to the lines in the image, a new unit vector $\boldsymbol{w}$, perpendicular to $\boldsymbol{v}$, is introduced to indicate the spacing between the lines. An expression for $\boldsymbol{w}$ can be obtained from the expression for $\boldsymbol{v}$ by replacing ϕ_z with $\phi_z + 90°$. When these three-dimensional vectors, $\boldsymbol{v}$ and $\boldsymbol{w}$, are projected onto an image plane, they generally do not remain perpendicular (e.g., $\boldsymbol{v}_l$ and $\boldsymbol{w}_l$ in Fig. 2B). If we let $\boldsymbol{v}_l^{\perp} = (-\boldsymbol{v}_{l_y}, \boldsymbol{v}_{l_x})$, 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, and a simple expression for the denominator can be found in terms of θ_l.

$$\boldsymbol{w}_l \cdot \boldsymbol{v}_l^{\perp} \;=\; \cos\phi_x \cos(\phi_y + \Delta\phi_y) \quad ; \quad ||\boldsymbol{v}_l|| \;=\; \left|\frac{\cos\phi_x \sin\phi_z}{\sin\theta_l}\right|$$

Combining these with similar expressions for λ_r gives a concise expression for spatial frequency disparity.

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

To determine spatial frequency disparity from a given pattern orientation ϕ_z, surface orientation ϕ_x, ϕ_y, and view angle $\Delta\phi_y$, this equation and the previous ones to determine θ_l, θ_r are all that are needed.

For solving the inverse problem (i.e., determining surface orientation), it has been shown that from the orientations θ_l, θ_r and θ_l', θ_r' of two corresponding line elements, (or θ_l, θ_r and λ_l, λ_r for parallel lines), the three-dimensional surface normal can be recovered (Jones, 1991). If more observations are available, they can be exploited using a least squares algorithm. This is based on the following expression:

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

This has the convenient interpretation that for a given surface orientation, all the observations (a_i, b_i) should lie along a straight line whose orientation gives ϕ_y and whose perpendicular distance from the origin is $\tan\phi_x$. Details of the derivation and experimental results may be found in (Jones and Malik, 1991).

In Section 4 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. 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.

3 Formulation using Gradient of Horizontal Disparity

Consider a region of a surface visible from two viewpoints. Let $P = (x, y)$ be the coordinates of a point within the projection of this region in one image, and $P' = (x', y')$ be the 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 — the image patch is one view is merely a translated version of its corresponding patch in the other view. 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_x, H_y, V_x, V_y specify the linear approximation to the distortion and are zero when the surface is fronto-parallel. 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 provides the best linear approximation. 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 and the vertical components of the image distortion are even smaller in practice. For this reason, it will be assumed that $V_x, V_y = 0$, leaving H_x which corresponds to a horizontal *compression* or *expansion*, and H_y which corresponds to a vertical *skew*. In both 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 (1980) 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:

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

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. For a planar surface, with orientation ϕ_x, ϕ_y, the image distortion is given by:

$$H_x = \frac{\cos(\phi_y - \Delta\phi_y)}{\cos(\phi_y + \Delta\phi_y)} - 1 \quad ; \quad H_y = \frac{\tan\phi_x \sin(2\Delta\phi_y)}{\cos(\phi_y + \Delta\phi_y)}$$

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 interchange the roles of the two viewpoints, or equivalently, the inverse of the transformation matrix can be computed directly. Compression and skew depend on the angular separation $2\Delta\phi_y$ of the viewpoints and are reduced as this angle decreases, since this is the angle subtended by the viewpoints, relative to a point on the surface. More distant surfaces lead to a smaller angle, making it more difficult to judge their inclination.

4 Surface Shape from Differences in Spatial Filter Outputs

We have been developing a new stereo algorithm based on the outputs of linear spatial filters at a range of orientations and scales. The collection of filter responses at a position in the image (the filter response vector, $v_R = F^T I_R$), provides a very rich description of the local image patch and can be used as the basis for establishing stereo correspondence (Jones and Malik, 1992). For slanted surfaces, however, even corresponding filter response vectors will differ, but in a way related to surface orientation. Such differences would normally be treated as noise in other stereo models.

From filter responses in the right image, we could, in principle, reconstruct the image patch using a linear transformation, namely the pseudo-inverse (for details and examples, Jones and Malik, 1992). For a particular surface slant, specified by H_x, H_y, we could predict what the image should look like in the other view, using another linear transformation — the affine transformation discussed earlier. A third linear transformation would predict the filter responses in the other view (Fig. 3).

$$v'_L = \underbrace{F^T \cdot T_{H_x,H_y} \cdot (F^T)^{-1}}_{M_{H_x,H_y}} v_R$$

Here the notation $(F^T)^{-1}$ denotes a pseudo-inverse. This sequence of transformations can, of course, be collapsed into a single one, M_{H_x,H_y}, that maps filter responses from one view directly to a prediction for filter responses in the other view. These M matrices depend on H_x and H_y but not on the input images, and can be pre-computed once, ahead of time. 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.

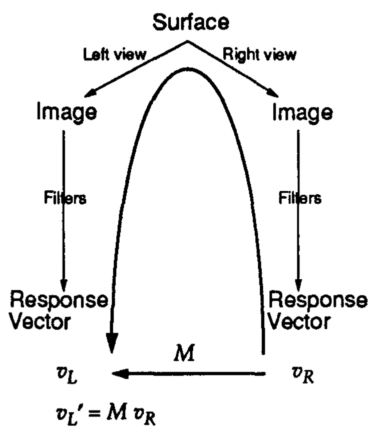

Fig. 3. Comparing spatial filter outputs to recover 3-D surface orientation.

This 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.

5 The Accuracy of Recovered Surface Orientations

This approach was tested quantitatively using randomly generated stereo pairs with known surface orientations (Fig. 4). For each of 49 test surface orientations ($H_x, H_y \in \{0.0, \pm 0.1, \pm 0.2, \pm 0.4\}$), 50 stereo pairs were created and the values of H_x, H_y were recovered using the method described in this paper. The recovered surface orientations (Fig. 5) are quite accurate, especially for small slants. For larger slants, the spread in the recovered surface orientation increases, similar to some psychophysical results. Small systematic errors, such as those for large rotations around the vertical axis, are likely not an inherent feature of the model, but an artifact of this particular implementation where surface orientation was computed from coarse estimates using the parabolic interpolation.

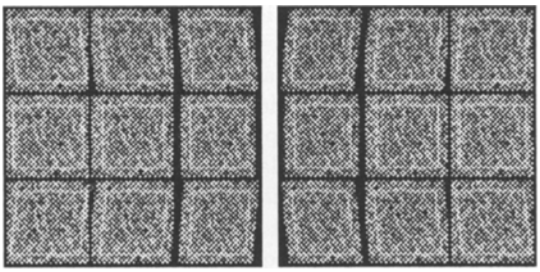

Fig. 4. Stereo pair of surfaces tilted in depth. A white square marked on each makes the horizontal compression/expansion, when $H_x \neq 0$, and vertical skew, when $H_y \neq 0$, quite apparent.

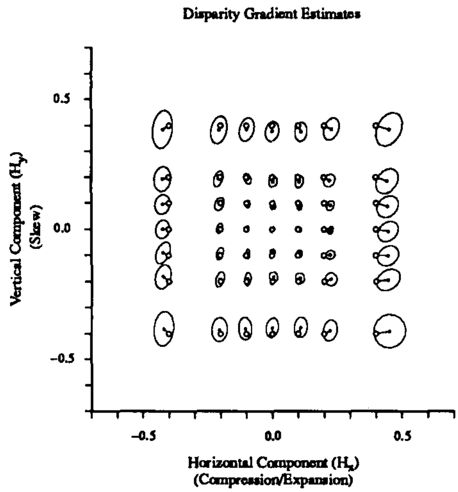

Fig. 5. Disparity gradient estimates. For various test surface orientations (*open circles*), the mean (*black dot*) and standard deviation (*ellipse*) of the recovered disparity gradient are shown.

Because the test surfaces are marked with random textures, the orientation and spatial frequency disparities at a single position encode surface orientation to varying degrees, and on some trials would provide only very limited cues. Horizontal stripes, for example, provide *no* information about a rotation around the vertical axis. For large planar surfaces, or smooth surfaces in general, estimates could be substantially refined by pooling over a local neighborhood, trading off spatial resolution for increased accuracy.

6 Orientation and Spatial Frequency Disparities Alone

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. However, orientation and spatial frequency disparities lead to the perception of a tilted surface, even in the *absence* of any systematic correspondence (Tyler and Sutter, 1979; von der Heydt et al., 1981). 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 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.

Given a pair of images, the implementation used in the previous experiment determines the best estimate of surface orientation — even if it is nonsense. This allow us to examine how it performs when the assumption of correspondence is false. Stereo pairs were created by filtering random, uncorrelated one-dimensional noise to have a bandwidth of 1.2 octaves and either an orientation disparity (Fig. 6A) or spatial frequency disparity. Since a different random seed is used for each image, there is no consistent correspondence or phase relationship. A sequence of 100 such pairs was created and for each, using 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.

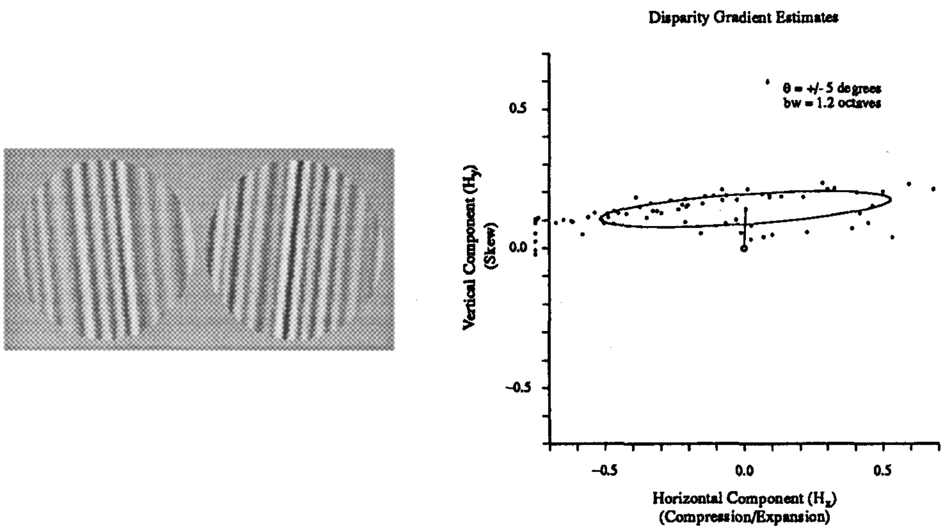

Fig. 6. *A*.Orientation disparity without correspondence. *B*. Disparity gradient estimates.

There is a fair bit of scatter in these estimates (Fig. 6B), but if the image pairs were presented rapidly, one after the other, 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 psychophysical 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. The same procedure was repeated for several different orientation disparities and for a considerable range, the recovered slant (H_y) increases with orientation disparity. Similar results were found for stereo pairs with spatial frequency disparities, but no systematic correspondence.

7 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

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

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

Jones DG, Malik J (1991) Determining three-dimensional shape from orientation and spatial frequency disparities I — using corresponding line elements. Technical Report UCB-CSD 91-656, University of California, Berkeley

Jones DG, Malik J (1992) A computational framework for determining stereo correspondence from a set of linear spatial filters. Proc ECCV Genova

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 IU 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

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 Science (suppl) 30:262

Tyler CW, Sutter EE (1979) Depth from spatial frequency difference: an old kind of stereopsis? Vision Research 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

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 1(2):133–144