Direct Recovery of Three-Dimensional Scene Geometry from Binocular Stereo Disparity

Richard P. Wildes, *Member, IEEE*

Abstract—**The problems under consideration center around the recovery of three-dimensional scene geometry from binocular stereo disparity. In order to accomplish this goal an analysis of disparity is presented. The analysis makes explicit the geometric relations between a stereo disparity field and a differentially projected scene. These results show how it is possible to recover three-dimensional surface geometry through first-order (i.e., distance and orientation of a surface relative to an observer) and binocular viewing parameters in a direct fashion from stereo disparity. As applications of the analysis, algorithms have been developed for recovering three-dimensional surface orientation and discontinuities from stereo disparity. The results of applying these algorithms to natural image binocular stereo disparity information are presented.**

Index Terms—**binocular stereo, disparity interpretation, three-dimentional vision.**

I. Introduction

A. Motivation

THE physical situation that leads to binocular stereo vision can be described as follows. An arrangement of surfaces in the three-dimensional world projects differentially onto a pair of two-dimensional imaging surfaces. To understand stereo vision would be to understand how the corresponding inverse mapping can take place. That is, given a pair of two-dimensional projections of a three-dimensional world, how is it possible to exploit the geometry of the situation to recover useful properties of the geometry of that world? At the current state of our understanding of stereo vision it is convenient to break the problem into three relatively independent parts: 1) feature selection, 2) correspondence, and 3) disparity interpretation. Feature selection is concerned with determining what makes for good primitive elements as one seeks to compare the pair of images in a binocular projection. Correspondence deals with matching those elements in the two views that are projections of the same element in the three-dimensional world. Disparity interpretation is concerned with how the disparity in configurations between corresponding elements can be mapped to useful descriptors of three-dimensional scene geometry. In this paper the focus of attention is the disparity interpretation problem.

Manuscript received September 3, 1989; revised March 1, 1991. Recommended for acceptance by J. L. Mundy. This work was supported by a graduate fellowship from the National Science Foundation and by the U.S. Department of Defense Advanced Research Projects Agency under Office of Naval Research Contract N00014-85-K-0124.

The author was with the Artificial Intelligence Laboratory and the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139. He is now with the S.R.I. David Sarnoff Research Center, Princeton, NJ 08543.

IEEE Log Number 9100182.

Typically, the disparity interpretation problem has been solved in terms of mapping stereo disparity into a dense representation of the distances between points in a scene and some reference point, perhaps the stereo viewer (see, e.g., Barnard and Fischler [2] for a discussion of this approach). However, consider the following question: Is the distance to the visible surfaces in a scene the only (or even the most) desirable output of stereo vision? As alternatives consider the possibilities of directly interpreting stereo disparity in terms of surface orientations and discontinuities as well as distance. Intuition suggests that information concerning these properties would be more useful to latter visually guided processes (e.g., manipulation, object recognition, and passive navigation) than would simple point by point distance. With these possibilities in mind, the research that is described in this paper takes a deeper look at the disparity interpretation problem.

B. Related Work

This subsection provides a brief review of related studies of interpreting stereo disparity. When useful, mention also will be made of studies in interpreting motion based disparity. For an extended review of approaches to interpreting visually based disparity see Wildes [40].

Perhaps the most popular approach to interpreting stereo disparity has been in terms of surface fitting. In its simplest form the idea behind the surface fitting approach is to interpolate (or approximate) the (possibly sparse) disparity values resulting from the correspondence process with a smooth surface. Technically the disparity values first should be converted to distance values; in practice the disparity values often are employed directly. The surface fitting idea has been instantiated in at least two forms: 1) minimization of spline functionals and 2) directly fitting polynomial based surface patches. The intuitive idea behind minimizing spline functionals is simple enough: Fit an elastic plate or membrane to the given data points and allow it to achieve equilibrium. The resulting representation is of point by point distance. The nontrivial technical details of applying this approach to disparity information have been the focus of much research (e.g., Grimson [10], Terzopoulos [33], Blake and Zisserman [3], and Boult [4]). The polynomial based approaches proceed by directly fitting a polynomial to the available distance data. Thus, the output representation is of an algebraic surface patch (e.g., Eastman and Waxman [9] and Hoff and Ahuja [13]).

In their most simple instantiations surface fitting approaches fail to make explicit the properties of surface orientation and discontinuity (in fact, discontinuities will tend to be smoothed

0162-8828/91/0600-0761$01.00 © 1991 IEEE

over). Various attempts have been made to extend the surface fitting approaches to deal with these matters. One tack on recovering surface orientation has centered around numerically differentiating the recovered point by point distance representations (Brady *et al.* [5], Medioni and Nevatia [25]). Other researchers have sought to recover surface orientation in the spline based approaches via a cascade of differencing operations that are in effect during the minimization process (Harris [12], Terzopoulos [34]). Also, the recovered coefficients of the surface patch methods often can be interpreted in terms of surface orientation or curvatures (Eastman and Waxman [9], Hoff and Ahuja [13]). The majority of approaches to recovering surface discontinuities within this paradigm are founded in either studies of the residuals of surface fitting (e.g., Grimson and Pavlidis [11], Eastman and Waxman [9], Hoff and Ahuja [13], and Lee and Pavlidis [20]) or nonconvex minimizations where discontinuities are allowed to form in a smooth surface at a penalty (e.g., Terzopoulos [33], Marroquin [23], Koch *et al.* [16], and Blake and Zisserman [3]).

A second approach to disparity interpretation can be found in differential imaging. Studies in differential imaging seek to understand the relation between scene geometry and an infinitesimal change of viewpoint. Analysis proceeds by first specifying a locally analytic form for a surface and then developing the difference equation for the surface's projection onto image planes related via an infinitesimal change of coordinates. The study of the resulting vector field can explicitly relate surface geometry (e.g., distance, orientation, and curvature with respect to the viewer) to the structure of projected disparity. This work has been done with reference to optical flow (e.g., Kanatani [15], Koenderink and van Doorn [18], Longuet-Higgins and Pradzny [21], Pradzny [27], Subbarao [32] and Waxman and Ullman [38]) as well as stereo vision (e.g., Eastman and Waxman [9], Koenderink and van Doorn [17], Mayhew and Longuet-Higgins [24], Weinshall [39]). Recently, it has been pointed out that similar work has been carried out for some time in the field of photogrammetry (Horn [14] and *Manual of Photogrammetry* [22]). It is worth noting that a common thread through many of these analyses is the application of tensor analysis to the classical field theory of mathematical physics (see Truesdell and Toupin [37]). While some differential imaging studies have explored the recovery of surface geometry of higher order than relative distance, little work on recovering discontinuities has been reported in this paradigm (excepting Eastman and Waxman [9]).

Finally, a number of studies have given particular attention to the recovery of surface discontinuities from disparity. Studies in both stereo and motion based disparity have suggested applying edge-detection methods to disparity in order to recover discontinuities (e.g., Clocksin [7], Thompson *et al.* [35], Schunk [28], and Stevens and Brookes [31]). Other studies have been based around the idea that the correlation between the images should be particularly low in the region corresponding to a discontinuity in the world (e.g., Smitley and Bajcsy [30] and Anandan [1]). Also, Mutch and Thompson [26] have analyzed occlusion as an aid in recovering discontinuities, as yet this method only has been applied to motion based disparity. Recently, Yeshurun and Schwartz [41] have exploited Cepstral filtering in an algorithm for recovering discontinuities in stereo disparity.

C. Contributions of this Paper

The research that is presented in the body of this paper bears some resemblance to several of the studies that have been reviewed. Most of the analytic developments presented in this paper are based in differential imaging. Therefore, the closest relatives to the presented work are found in earlier studies of differential imaging, in particular Koenderink and van Doorn [17], [18]. However, the current work makes a number of novel contributions to the disparity interpretation problem. The most significant points of distinction are as follows.

- The recovery of three-dimensional surface geometry (i.e., orientation and discontinuities, in addition to relative distance) emphasizes operations that proceed directly from stereo disparity. This tack is in contrast to many standard approaches where higher order surface geometry is derived only indirectly from distance information.
- Novel relations between the differentially projected orientation of surface detail (e.g., texture) and underlying three-dimensional surface geometry are presented. These relations allow three-dimensional surface orientation, distance, and binocular viewing parameters to be specified explicitly in terms of stereoscopic disparity.
- An extensive stability analysis is presented for the equations that relate three-dimensional surface orientation, distance, and binocular viewing parameters to stereo disparity. The results of the stability analysis indicate not only the requirements for the accurate recovery of surface geometry and viewing geometry, but also how disparity interpretation algorithms can monitor the reliability of their own output.
- The analysis of stereo disparity that is developed leads to algorithms for recovering three-dimensional surface orientation and discontinuities from stereo disparity. These algorithms have been implemented and tested on natural image stereo disparity data.

D. Outline of Paper

The remainder of this paper unfolds in four sections. In Section II a formal analysis of stereo disparity is presented. The analysis makes explicit the relation between the geometry of stereo disparity and a differentially viewed world. Section III presents a stability analyses of the relations between stereo disparity measures and three-dimensional scene geometry, as derived in Section II. In Section IV the results of the previous two sections are used to motivate algorithms for recovering three-dimensional surface orientations and discontinuities from stereo disparity. The results of applying these algorithms to stereo disparity also are presented. Finally, Section V provides a summary.

II. Formal Developments

In this section a formal analysis of stereo disparity is presented. The goal is to explicitly relate disparity based measurements to differentially projected scene geometry. There-

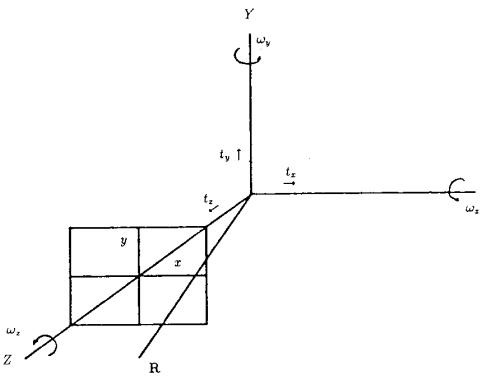


Fig. 1. A general infinitesimal change of coordinates is composed of a translation $\mathbf{T} = (t_x, t_y, t_z)$ and a rotation $\mathbf{\Omega} = (\omega_x, \omega_y, \omega_z)$. A point $\mathbf{R} = (X, Y, Z)$ undergoes perspective projection onto a plane located at $Z = 1$.

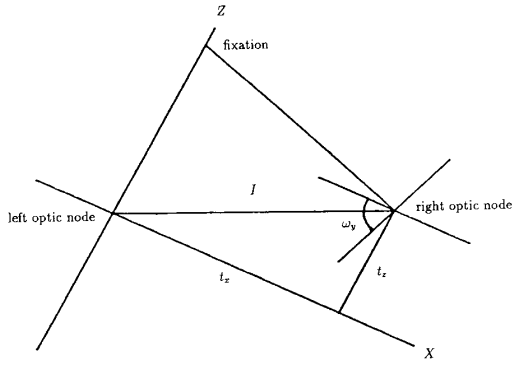


Fig. 2. A model of binocular stereo viewing geometry. Two views fixate a point in space. The views are represented by their optical nodes and are separated by a stereo baseline of length I. A coordinate system is defined at the left optical node. The right optical node is related to the left by translations along the X and Z axes and by rotation about the Y-axis.

fore, the analysis is based in understanding the process of differentially projecting a three-dimensional world onto a pair of two-dimensional imaging surfaces. In turn, attention will be given to the forward mapping of the world into the image(s) and the inverse mapping that allows for recovery of three-dimensional scene geometry from stereo disparity.

A. Analysis of Disparity

Given a general change in coordinate systems the corresponding change to a point $\mathbf{R}$ can be described as

$$\delta\mathbf{R} = -\mathbf{T} - (\mathbf{\Omega} \times \mathbf{R}) \tag{1}$$

where the symbols are described with reference to Fig. 1.[1]

Now, for the case of binocular stereo vision it is not necessary to deal with the most general change of coordinates as described by (1). Instead, consideration can be restricted to the model of stereo geometry as given in Fig. 2.

This system is related to a coordinate system defined at the optical center of the left view. The translation components are confined to the plane defined by the view direction and the axis connecting the two optical nodes; therefore, $t_y = 0$. The rotation is confined to rotation about the y-axis; therefore, $\omega_x = \omega_z = 0$. This is not to say that elevation of the views is not permitted. Rather, the coordinate system is simply always moved with the elevation. For this situation, substitution into (1) yields

$$\delta\mathbf{R} = -(t_x + \omega_y Z, 0, t_z - \omega_y X). \tag{2}$$

Perspective projection serves as the model of how the world projects into an image plane. The laws of perspective yield (with appropriate units)

$$x = \frac{X}{Z} \quad y = \frac{Y}{Z}. \tag{3}$$

[1]A few comments on notation: Throughout this presentation bold font is used for vectors. Uppercase letters X, Y, and Z denote world coordinates; while lowercase x and y denote image coordinates. Subscripts are used for vector components, not to denote differentiation.

To understand how a point in space changes its Euclidean image coordinates under a projective transformation from one view to another let disparity be defined as

$$\boldsymbol{\chi} = (\chi_x, \chi_y) = (\delta x, \delta y). \tag{4}$$

Considering the model of projection embodied in (3) it is found that

$$\boldsymbol{\chi} = \left(\frac{\delta X}{Z} - X\frac{\delta Z}{Z^2}, \frac{\delta Y}{Z} - Y\frac{\delta Z}{Z^2}\right). \tag{5}$$

Then upon substituting (2) into (5)

$$\boldsymbol{\chi} = \left(\frac{1}{Z}(xt_z - t_x) - (x^2+1)\omega_y, \frac{1}{Z}(yt_z) - xy\omega_y\right). \tag{6}$$

The equations in (6) present the basic relations for stereo disparity.

In order to gain further relations between the variables, and in particular to gain relations that concern the gradient of distance (which later will allow the recovery of surface orientation) it is useful to consider the gradient of disparity. This gradient is a first-order tensor of the following form

$$\chi' = \begin{pmatrix} \frac{\partial \chi_x}{\partial x} & \frac{\partial \chi_x}{\partial y} \\ \frac{\partial \chi_y}{\partial x} & \frac{\partial \chi_y}{\partial y} \end{pmatrix} \tag{7}$$

where

$$\begin{aligned} \frac{\partial \chi_x}{\partial x} &= \frac{t_z}{Z} + \left(\frac{\partial}{\partial x}\frac{1}{Z}\right)(xt_z - t_x) - 2\omega_y x \\ \frac{\partial \chi_x}{\partial y} &= \left(\frac{\partial}{\partial y}\frac{1}{Z}\right)(xt_z - t_x) \\ \frac{\partial \chi_y}{\partial x} &= \left(\frac{\partial}{\partial x}\frac{1}{Z}\right)(yt_z) - \omega_y y \\ \frac{\partial \chi_y}{\partial y} &= \frac{t_z}{Z} - \left(\frac{\partial}{\partial y}\frac{1}{Z}\right)(yt_z) - \omega_y x. \end{aligned} \tag{8}$$

To further interpret (8) it is useful to decide upon a representation for the distance parameter Z. Suppose that consideration

is limited to representing surface geometry only through first-order. In this case, consider the standard first-order surface representation

$$Z = pX + qY + r \tag{9}$$

where $(p, q) = \nabla Z$ is the surface gradient and r is the distance along the Z-axis. In terms of image coordinates, (x, y), (9) becomes

$$\frac{1}{Z} = \frac{1 - px - qy}{r} \tag{10}$$

and therefore

$$\frac{\partial}{\partial x}\left(\frac{1}{Z}\right) = -\frac{p}{r}, \frac{\partial}{\partial y}\left(\frac{1}{Z}\right) = -\frac{q}{r}. \tag{11}$$

Upon substituting (11) into (8) and retaining terms through first-order, it is found that

$$\begin{aligned}
\frac{\partial \chi_x}{\partial x} &= \frac{t_z}{Z} - \frac{p}{r}(xt_z - t_x) - 2\omega_y x \\
\frac{\partial \chi_x}{\partial y} &= -\frac{q}{r}(xt_z - t_x) \\
\frac{\partial \chi_y}{\partial x} &= -\frac{p}{r}(yt_z) - \omega_y y \\
\frac{\partial \chi_y}{\partial y} &= \frac{t_z}{Z} - \frac{q}{r}(yt_z) - \omega_y x.
\end{aligned} \tag{12}$$

The representation of χ' can be simplified further by choosing the coordinate system in an appropriate fashion. In particular, let the coordinate system be oriented so that the Z-axis is pointing along the line of regard. The appropriate change of coordinates is given in terms of Euler angles by transforming the original system according to

$$\begin{pmatrix} \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \\ \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \end{pmatrix} \tag{13}$$

where θ and ϕ are the spherical polar coordinates of the point of regard; see Korn and Korn [19]. In this system $x = y = 0$ while $Z = r$. Therefore, the gradient tensor of disparity can be written as

$$\chi' = \begin{pmatrix} \frac{1}{r}(pt_x + t_z) & \frac{q}{r}t_x \\ 0 & \frac{t_z}{r} \end{pmatrix}. \tag{14}$$

Recalling that an eventual goal is the recovery of useful relations involving geometric surface parameters, such as orientation (p, q) and distance r , it is pleasing to see these terms appear in the final form of χ' given in (14).

For purposes of analysis it is convenient to split χ' into its symmetric, χ'_+, and antisymmetric, χ'_-, parts. This yields

$$\chi' = \chi'_+ + \chi'_- = \frac{1}{2}\begin{pmatrix} \frac{2}{r}(pt_x + t_z) & \frac{q}{r}t_x \\ \frac{q}{r}t_x & 2\frac{t_z}{r} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & \frac{q}{r}t_x \\ -\frac{q}{r}t_x & 0 \end{pmatrix}.$$

Physically, χ'_+ describes the nonrigid change in shape as an object is differentially projected; while χ'_- describes how an object is rigidly rotated through differential imaging. This interpretation follows directly from the Cauchy–Stokes decomposition theorem of tensor analysis (Truesdell and Toupin [37]). For most of the rest of this paper, attention will be restricted to the properties of χ'_+ as it has proven to give the most insight into interpreting the disparity field.

In order to understand the nature of χ'_+ it is useful to analyze the structure of its eigenvalues and eigenvectors. (Intuitively speaking, this analysis will yield information about the direction and magnitude of the nonrigid transformation embodied in χ'_+.) The characteristic equation, $\det(\chi'_+ - \lambda \mathrm{I}) = 0$ (where $\mathbf{I}$ is the identity matrix), corresponding to χ'_+ is

$$\lambda^2 - \frac{1}{r}(pt_x + 2t_z)\lambda + \frac{1}{r^2}[pt_x t_z + t_z^2 - \frac{(qt_x)^2}{4}] = 0$$

the roots of which, and hence the eigenvalues, are

$$\lambda = \frac{1}{2r}[pt_x + 2t_z \pm (p^2 + q^2)^{\frac{1}{2}}]. \tag{15}$$

For each eigenvalue, λ_i, the equation $(\chi'_+ - \lambda_i \mathrm{I})\boldsymbol{\xi}_i = 0$ yields the corresponding eigenvector ξ_i . This yields

$$\boldsymbol{\xi} = \frac{\left(p + (p^2 + q^2)^{\frac{1}{2}}, q\right)}{(p^2 + q^2)^{\frac{1}{2}}} \tag{16}$$

as the eigenvector corresponding to the positive root of (15). The eigenvector corresponding to the negative root is found to be orthogonal to (16). The standard interpretation of such results says that χ'_+ operates on an object by stretching it in amount λ_i along the direction specified by $\boldsymbol{\xi}_i$.

Should the two values assigned to λ by (15) be unequal the deformation embodied by χ'_+ is nonisotropic. To make this notion precise define

$$\sigma = \lambda_{\max} - \lambda_{\min} = \frac{t_x}{r}(p^2 + q^2)^{\frac{1}{2}}. \tag{17}$$

Physically, σ accounts for an area-preserving, but nonconformal transformation between differentially projected images. It may be interpreted as a contraction along the direction of one of the eigenvectors, (16), with a corresponding expansion along the other eigenvector. Most interesting for present concerns are the following two observations. First, the value for σ, the contraction/expansion, as specified by (17) is the product of the magnitude of the surface gradient, $(p^2 + q^2)^{\frac{1}{2}}$, and the distance scaled view translation orthogonal to the Z-axis, $\frac{t_x}{r}$. Second, the ray defined by $\boldsymbol{\xi}$, the axis of contraction, as specified by (16) is halfway between the rays defined by the surface gradient, (p, q), and $(t_x, 0)$, the component of translation orthogonal to the Z-axis.

Now, consider the following intuition. In so far as σ is a nonconformal transformation, it should be possible to relate it explicitly to a change in how angles appear in the differential projections. This clearly would be a desirable result as a change in angles should be measurable directly from a pair of images. Now, the change in orientation of a linear segment is due to the operation of χ'. To understand this it is helpful to express χ' as

$$\chi' = \frac{1}{2}\begin{pmatrix} 0 & \frac{q}{r}t_x \\ -\frac{q}{r}t_x & 0 \end{pmatrix} + \frac{1}{2}\left(\hat{\xi}_1, \hat{\xi}_2\right)\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\left(\hat{\xi}_1, \hat{\xi}_2\right)^{-1}$$

where $\hat{\boldsymbol{\xi}}_i$ are the normalized eigenvectors. Next, suppose that the normalized eigenvectors are represented in terms of an

angle θ that represents their orientation with reference to the image coordinates. Then $2\chi'$ can be written as

$$\begin{pmatrix} 0 & \frac{q}{r}t_x \\ -\frac{q}{r}t_x & 0 \end{pmatrix} + \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{18}$$

Now, to obtain a relation concerning how a linear segment changes orientation between views: First, apply (18) to an oriented segment $(\cos\psi, \sin\psi)$. Second, take the cross-product of the result with the same oriented segment. After some amount of algebraic manipulation it is found that to first-order the sine of the angle between the initial segment and the transformed segment is

$$\frac{1}{2}[\sigma \sin 2(\psi - \theta) - qt_x]. \tag{19}$$

By taking the difference of two such measurements, that is a difference in projected angles, the effects of the rigid rotation, qt_x, are discounted. Thus, the suspicion that a change in angles mediated by differential projection should directly reflect the effects of σ is confirmed.

Finally, consider the relation of the vector quantities $\boldsymbol{\xi}_i$ to disparity-based measurements. Following through on the difference of two orientational disparities as defined in (19) yields

$$\frac{\sigma}{2}[\cos\theta(\sin\psi_1 - \sin\psi_2) + \sin\theta(\cos\psi_2 - \cos\psi_1)] \tag{20}$$

where as before $(\cos\theta, \sin\theta)$ specifies the direction of the axes $\boldsymbol{\xi}_i$ and the $(\cos\psi_j, \sin\psi_j)$, $j = 1, 2$, specify the directions of the two differentially projected oriented segments. Notice that only the directions of the $\boldsymbol{\xi}_i$ are important. Therefore, an additional pair of orientational disparities allows the unique determination of the eigenvectors $\boldsymbol{\xi}_i$.

The results just presented concerning the gradient of disparity have concentrated on the case where the coordinate system has been chosen so that the Z-axis (the optical axis of one of the imaging systems) is oriented along the line of regard. Interestingly, matters are similar for less fortuitous choices of coordinate system. To understand this, the eigenvalues and eigenvectors of the gradient of disparity, χ', and in particular its symmetric component, χ'_+, can be calculated using the terms specified by (12), rather than the simplified forms given in (14). In this case, to first-order in image coordinates, the contraction/expansion operation is given by

$$\sigma = \frac{\|\mathbf{T_v}\|}{r_\mathbf{v}}(p_\mathbf{v}^2 + q_\mathbf{v}^2)^{\frac{1}{2}},$$

while the axis of contraction becomes

$$\boldsymbol{\xi} = \frac{\|\mathbf{T_v}\|(p_\mathbf{v}, q_\mathbf{v}) + \|(p_\mathbf{v}, q_\mathbf{v})\|\mathbf{T_v}}{\|\|\mathbf{T_v}\|(p_\mathbf{v}, q_\mathbf{v}) + \|(p_\mathbf{v}, q_\mathbf{v})\|\mathbf{T_v}\|}$$

Here, for a visual direction $\mathbf{v}$ defined with respect to image coordinates (x,y) : $\mathbf{T_v}$ is the component of $\mathbf{T}$ orthogonal to $\mathbf{v}$; $r_\mathbf{v}$ is the distance to the viewed surface along $\mathbf{v}$; $(p_\mathbf{v}, q_\mathbf{v})$ is the gradient of the viewed surface with respect to $\mathbf{v}$. These results are a straightforward generalization of the interpretations of σ and $\boldsymbol{\xi}$ embodied in equations (17) and (16), respectively.

Recapitulation: This subsection has developed an analysis of stereo disparity. In following developments the key results attained so far will be (6), (16), (17), (19), and (20). Notice that relations that involve three different types of disparity information have been derived: horizontal disparity (6), vertical disparity (6), and orientational disparity (16), (17), (19) and (20). These relations will yield straightforward relations between stereo disparity, stereo viewing parameters and geometric surface parameters p, q, and r.

B. Recovery of Scene Geometry

At this point an analysis is in hand that relates stereo disparity to scene geometry. In this subsection the analysis is exploited to show how three-dimensional scene geometry can be recovered from disparity. The first part of this discussion shows how geometric surface parameters representing distance, r, and surface orientation, (p, q), can be recovered. The second part shows how the parameters that relate the pair of images comprising a binocular view, t_x, t_z and ω_y, can be recovered.

1) Recovering Geometric Surface Parameters: Given the developed analysis of stereo disparity, relations for recovering geometric surface parameters, p, q and r, are obtained easily. To begin, consider recovery of the distance parameter, r. Provided the differential viewing parameters are known, the Z value of any point can be recovered via a measure of either horizontal or vertical disparity and the corresponding relation from (6). By adopting a coordinate system where the Z-axis is toward the point of observation (as discussed in Section II-A) the recovered Z value can be interpreted as r. Typically, horizontal disparity is the measure of choice for recovering distance in this situation. For most binocular viewing conditions horizontal disparity has considerably greater magnitude than vertical disparity and therefore can be measured more accurately.

Now, consider the recovery of surface orientation, (p, q). Recall that (16) implies that the ray defined by $\boldsymbol{\xi}_1$ is halfway between the rays defined by (p, q) and the X-axis. This observation leads one to note that

$$(1, 0) \cdot \frac{(p, q)}{(p^2 + q^2)^{\frac{1}{2}}} = \hat{\boldsymbol{\xi}}_{1x}^2 - \hat{\boldsymbol{\xi}}_{1y}^2$$

and

$$(1, 0) \times \frac{(p, q)}{(p^2 + q^2)^{\frac{1}{2}}} = 2\hat{\boldsymbol{\xi}}_{1x}\hat{\boldsymbol{\xi}}_{1y}$$

or

$$\frac{(p, q)}{(p^2 + q^2)^{\frac{1}{2}}} = (\hat{\boldsymbol{\xi}}_{1x}^2 - \hat{\boldsymbol{\xi}}_{1y}^2, 2\hat{\boldsymbol{\xi}}_{1x}\hat{\boldsymbol{\xi}}_{1y}) \tag{21}$$

where $\hat{\boldsymbol{\xi}}_1 = (\hat{\xi}_{1x}, \hat{\xi}_{1y})$ is $\boldsymbol{\xi}_1$ normalized. Now, rewriting (17) as

$$(p^2 + q^2)^{\frac{1}{2}} = \frac{r\sigma}{t_x}$$

allows substitution into (21) for the term $(p^2 + q^2)^{\frac{1}{2}}$ with the result that the surface orientation parameters, p and q, can be

recovered as

$$p = (\hat{\xi}_{1x}^2 - \hat{\xi}_{1y}^2)\frac{r\sigma}{t_x}$$
$$q = (2\hat{\xi}_{1x}\hat{\xi}_{1y})\frac{r\sigma}{t_x} \quad . \tag{22}$$

The relations specified in (22) allow for the recovery of surface orientation, (p, q), up to a scale factor, $\frac{r}{t_x}$, in terms of quantities, σ and $\boldsymbol{\xi}$, that are recoverable from a disparity map. If the binocular viewing geometry is known, then the surface orientation can be recovered exactly. In this case, t_x is known while r can be recovered from a measure of horizontal disparity.

Relations (22) specify surface orientation in terms of the parameters of stereoscopic expansion/contraction, σ and $\boldsymbol{\xi}$. Earlier, (20) was derived to relate σ and ξ to orientational disparity. Therefore, it is interesting to develop the direct relationship between surface orientation and orientational disparity. In particular, appropriate substitution of the terms p, q, r, and t_x into (20) followed by algebraic manipulation yields

$$\frac{t_x}{r}\begin{pmatrix} \sin 2\psi_1 - \sin 2\psi_2 & \cos 2\psi_1 - \cos 2\psi_2 \\ \sin 2\psi_3 - \sin 2\psi_4 & \cos 2\psi_3 - \cos 2\psi_4 \end{pmatrix}\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} \Delta_{1,2} \\ \Delta_{3,4} \end{pmatrix}$$

where $(\cos\psi_i, \sin\psi_i)$ is an oriented image segment and $\Delta_{i,i+1}$ is the difference in orientational disparities of two oriented image segments i and $i+1$.

2) Recovering the Differential View: In the event that the differential viewing parameters t_x, t_z and ω_y are not known in advance it is possible to exploit the analysis of stereo disparity to allow for their recovery. The formulation exploits horizontal and orientational disparity. (Vertical disparity is eschewed for reasons of accuracy as discussed above.) The presented method works with the assumption that the magnitude of the stereo baseline is a known value, say I. In the end, the method recovers the viewing parameters only up to an arbitrary distance scaling factor. This is due to the fact that the distance to some point in the world is assigned an arbitrary value in the course of the solution.

To begin these developments, substitute (10) into the horizontal disparity relation from (6) to obtain

$$\chi_x = \left(\frac{1 - px - qy}{r}\right)(xt_z - t_x) - (x^2 + 1)\omega_y. \tag{23}$$

Now, notice that at the mutual fixation point of the two views, $(x, y) = (0, 0)$, (23) reduces to

$$0 = \frac{1}{r}(-t_x) - \omega_y$$

or

$$\omega_y = \frac{-t_x}{r}. \tag{24}$$

Substitution of (24) into (23) allows for the elimination of one of the view parameters, ω_y.

The next step is to use orientational disparity to eliminate the surface orientation parameters p and q from (23). This goal is accomplished by exploiting the expressions for p and q given in (22).

Now, the viewing parameters t_x and t_z can be related in a relatively simple equation with one further manipulation:

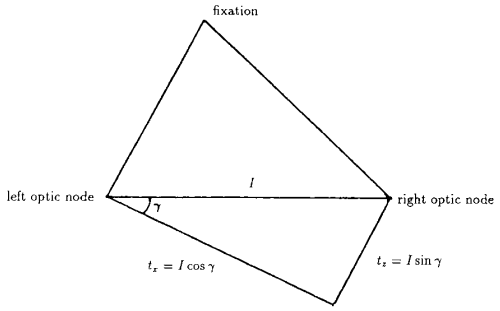


Fig. 3. The analysis of stereo disparity can be aided by making several geometric substitutions. To accomplish this γ is defined as the angle between the stereo baseline I and the X translation t_x.

Allow for an arbitrary distance scale and set the remaining surface parameter, r, to an arbitrary value of unity. This yields a relation of the form

$$a_0 = a_1 t_x + a_2 t_z + a_3 \frac{t_x}{t_z} \tag{25}$$

where the a_i consist entirely of known (or measurable) quantities. Explicitly,

$$a_0 = \chi_x - \sigma[x(\hat{\xi}_{1x}^2 - \hat{\xi}_{1y}^2) + 2y\hat{\xi}_{1x}\hat{\xi}_{1y}]$$
$$a_1 = x^2$$
$$a_2 = x$$
$$a_3 = -\sigma[x^2(\hat{\xi}_{1x}^2 - \hat{\xi}_{1y}^2) + 2xy\hat{\xi}_{1x}\hat{\xi}_{1y}].$$

Relation (25) could be used to solve for the desired parameters t_x and t_z in a number of ways. Here, the system is solved by making use of several substitutions and a small angle approximation. Let γ be defined as shown in Fig. 3. From Fig. 3 it is clear that

$$t_x = I\cos\gamma$$
$$t_z = I\sin\gamma$$
$$\frac{t_z}{t_x} = \tan\gamma. \tag{26}$$

Substituting (26) into (25) results in

$$a_0 = a_1 I\cos\gamma + a_2 I\sin\gamma + a_3\tan\gamma. \tag{27}$$

The next step is to take standard first-order trigonometric substitutions so that γ may be solved for as

$$\gamma = \frac{a_0 - a_1 I}{a_3 + a_2 I} \tag{28}$$

with I known. (The assumption that γ is small amounts to the conditions that the fixation is at a moderate distance and not overly eccentric. These conditions are in accord with the vast majority of real world viewing situations.) With γ recovered the original view parameters t_x, t_z and ω_y are easily obtained with reference to (26) and (24).

3) Recapitulation: The second half of this section has shown how the analysis of stereo disparity can be used to

locally invert the disparity map. In particular, relations have been presented that allow for the recovery of stereo viewing parameters (24), (26), and (28) surface orientation (22) as well as surface distance. Significantly, the recovery of these various parameters can proceed directly from measurements of horizontal and orientational disparity.

III. Stability Analysis

At this point it is useful to analyze the numerical properties of the proposed recovery methods. In turn, this section considers degenerate sets of measurements and sensitivity to measurement errors. The section closes with a recapitulation of its main results.

A. Degeneracies

It is possible that certain combinations of measured disparities and image coordinates will lead to situations where the proposed recovery methods will be undefined. This subsection presents an analysis of these degenerate situations. Of particular interest will be combinations of data that lead to a ratio becoming undefined as its denominator tends to zero.

Consider first the key relation for defining the viewing parameters, (28). Relation (28) will become undefined as its denominator approaches zero. Therefore, it is necessary to consider the condition

$$0 = a_3 + a_2 I$$

or, upon appropriate substitution

$$0 = xI - \sigma[x^2(\hat{\xi}_{1x}^2 - \hat{\xi}_{1y}^2) + 2xy\hat{\xi}_{1x}\hat{\xi}_{1y}].$$

Examination of this quantity indicates that the image line $x = 0$ is degenerate. Continuing by making the substitutions implied by (17) and (21) and cancelling appropriately yields

$$0 = I - \frac{t_x}{r}(px + qy)$$

or

$$1 = \frac{t_x(px + qy)}{Ir} \tag{29}$$

as a degenerate condition. In words, the numerator of (29) is the product of two factors: the first factor, t_x, is the projection of the stereo baseline on the X-axis. The second factor, $(px+qy)$, is the radial distance from the point of regard to the Z-intercept of the corresponding plane. The denominator of (29) is the product of the stereo baseline magnitude and the Z-intercept of the surface of regard. These two quantities must be equal for the viewing solution (28) to be undefined. It is quite unlikely that such a configuration will occur generically. For intuition, notice that in typical viewing conditions $\|t_x\| \approx \|I\|$. Therefore, the degenerate condition demands that the surface of regard is viewed at a point where it is approximately the same radial distance from its Z-intercept as the Z-intercept is from the viewer. See Fig. 4.

Now, turn attention to degeneracies related to the recovery of the components of the surface gradient $\nabla Z = (p, q)$ defined in (22). Two conditions present themselves. First, should the plane of consideration pass through the origin (i.e., the optical center of the left view) the solution will not apply. In this situation the plane appears as a line to the left view. Second, should $t_x = 0$ then (22) is undefined. For binocular stereo vision this a mechanical impossibility as it requires one optical node to be directly behind (and hence see through) the other optical node. Happily, neither of these degeneracies of the surface orientation recovery methods are likely to occur generically.

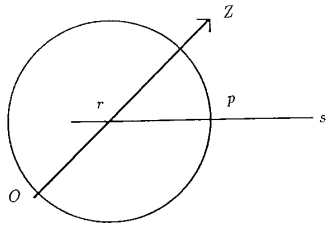


Fig. 4. A geometric configuration of surface and viewer leading to a degeneracy of the proposed method for recovering view and surface geometry. An observer o views a point p on a planar surface S. Suppose that S intercepts the Z-axis at r. The degenerate condition is $\|or\| = \|pr\|$; the points o and p must lie on a circle centered at r.

In general, this discussion leads to two positive conclusions concerning the degeneracies of the proposed recovery methods: First, while there is a degeneracy of practical importance for the view recovery methods (the image line $x = 0$), this condition can be easily diagnosed during the recovery process. Second, the remaining degenerate conditions for view and surface geometry recovery are either unlikely to occur or impossible for a binocular stereo system.

B. Error Analysis

Now consider the effects of applying error perturbations to the data that serves as input to the recovery methods, i.e., the disparity measures χ_x, σ and $\boldsymbol{\xi}$.[2] In particular, the goal of this section is to derive those conditions and choices of image measurements that lead to especially stable or unstable solutions. As a measure of stability the "generalized error-propagation formula" (Dahlquist and Björk [8]) will be used. This measure, which tells the local rate of change of a solution with respect to perturbed data, can be written

$$\Delta y \approx \sum_{i=1}^{n} \frac{\partial y}{\partial x_i}(\tilde{\mathbf{x}}) \cdot \Delta x_i$$

therefore

$$\|\Delta y\| \lesssim \sum_{i=1}^{n} \|\frac{\partial y}{\partial x_i}(\tilde{\mathbf{x}})\| \cdot \|\Delta x_i\| \tag{30}$$

where

$$y = y(\mathbf{x})$$

[2]To make the error analysis manageable it is assumed that errors in the assessment of image positions [i.e., (x,y)] are negligible as compared to those in disparity measurements. Therefore, the following developments will consider only perturbations to the disparity measurements.

is a solution based on data

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)$$

and the perturbation to $x_i is \Delta x_i$, resulting in

$$\tilde{x_i} = x_i + \Delta x_i$$
$$\tilde{\mathbf{x}} = (\tilde{x_1}, \tilde{x_2}, \ldots, \tilde{x_n})$$
$$\Delta y = y(\tilde{\mathbf{x}}) - y(\mathbf{x}).$$

Clearly, small values for $\|\Delta y\|$ correspond to stable solutions. The remainder of this subsection exploits the "generalized error-propagation formula'' to examine in turn the recovery methods for stereo viewing parameters, surface distance and surface orientation. As the discussion proceeds, the analytic conditions for stability will be explained in terms of both image and scene geometry.

First, turn attention to the stability of the view recovery method (28). Then the parameters of the generalized error-propagation formula, (30), become $\mathbf{x} = (\chi_x, \sigma, \theta)$ and $\tilde{\mathbf{x}} = \mathbf{x} + (\Delta\chi_x, \Delta\sigma, \Delta\theta)$, with $(\cos\theta, \sin\theta)$ specifying the direction of $\boldsymbol{\xi}$. Considering (30), the goal is to understand the conditions where

$$\|\frac{\partial\gamma}{\partial\chi_x}(\tilde{\mathbf{x}})\| \cdot \|\Delta\chi_x\| + \|\frac{\partial\gamma}{\partial\sigma}(\tilde{\mathbf{x}})\| \cdot \|\Delta\sigma\| + \|\frac{\partial\gamma}{\partial\theta}(\tilde{\mathbf{x}})\| \cdot \|\Delta\theta\| \quad (31)$$

is small. To begin, notice that trivially (31) can have arbitrarily small magnitude as $(\Delta\chi_x, \Delta\sigma, \Delta\theta) \rightarrow (0,0,0)$. More realistically, the perturbations, $(\Delta\chi_x, \Delta\sigma, \Delta\theta)$, need to be small compared to the denominators of the corresponding partial derivatives. These denominators will now be examined in some detail. The term $\frac{\partial\gamma}{\partial\chi_x}(\tilde{\mathbf{x}})$ can be expanded (with the aid of double angle formulas) as

$$-x[\tilde{\sigma}(x,y) \cdot (\cos 2\theta, \sin 2\theta) + I]^{-1}$$

or

$$-x[\tilde{\sigma}\|(x,y)\| \cos\psi + I]^{-1} \quad (32)$$

where ψ is the angle between (x,y) and $(\cos 2\tilde{\theta}, \sin 2\tilde{\theta})$. From (32) it can be concluded that the error due to the first term of (31) can be made small given three conditions: 1) σ, the magnitude of stereoscopic contraction/expansion, is large; 2) I, the magnitude of the stereo baseline, is large; 3) (x,y) is chosen in the direction of twice θ (i.e., twice the directions of the σ-axis, ξ). Using similar procedures and nomenclature, the second term of (31) can be written

$$\frac{\|(x,y)\| \cos\psi[\tilde{\chi_x} - I(1+x^2) - 2\tilde{\sigma}\|(x,y)\| \cos\psi]}{x(\tilde{\sigma}\|(x,y)\| \cos\psi + I)^2}. \quad (33)$$

By inspection it is possible to conclude that (33) has small magnitude when σ and I are relatively large. In the limit, as $\|(x,y)\| \rightarrow \infty$ l'Hopital's rule (Korn and Korn [19]) suggests that taking (x,y) in the direction 2θ is useful provided that σ is relatively large. For more moderate $\|(x,y)\|$ case analyses still suggest this is the appropriate direction for (x,y). The last term of (31) can be written

$$\frac{2\tilde{\sigma}\|(x,y)\| \sin\psi[\tilde{\chi_x} - I(1+x^2)] - 4\tilde{\sigma}\|(x,y)\|^2 \sin 2\psi}{x[\tilde{\sigma}\|(x,y)\| \cos\psi + I]^2}. \quad (34)$$

Consideration of (34) reveals that σ and I large with (x,y) chosen in the direction 2θ leads to its small magnitude. It also is useful for $\chi_x \approx -I(1+x^2)$. Notice in particular that the numerator of (34) contains terms of $\sin\psi$. This heightens the importance of choosing (x,y) in the direction 2θ.

Stepping back from the analysis allows for the observation that there are essentially three conditions for stability in the view recovery. First, the magnitude of stereoscopic contraction/expansion, σ, should be relatively large. In terms of scene geometry this condition implies that the magnitude of the surface gradient is large while the viewing distance is not too great. This is just another way of saying that the differential perspective information must be salient. Second, the image coordinates, (x,y), should be chosen in the direction of twice θ (i.e., twice the directions of the σ-axis, ξ). In the scene this means that the image coordinates should be chosen in the direction of ∇Z, see equation (21). Intuitively, the data points are best when chosen in the direction of the projection of the surface gradient. Third, the magnitude of the stereo baseline, I, should be relatively large. Again, this condition is related to making the disparity information as salient as possible. It is interesting to note that the third condition can be satisfied as a one time design constraint on a stereo system. Similarly, notice that the first two conditions can be monitored by an intelligent disparity processing algorithm. This last observation deserves emphasis. The error analysis indicates that the view recovery method can guide its own behavior in order to recover a stable solution given input disparity information.

Now, consider how errors in the measurements affect the recovery of the local distance parameter, r. By an appropriate local transformation [e.g., the coordinate transformation matrix (13)] r is recovered in terms of Z. Therefore, the appropriate relation is (6). For the following analysis the potential sources of error will be in the horizontal disparity, χ_x, as well as the previously recovered view parameters t_x, t_z and ω_y. Therefore, the parameters of the generalized error-propagation formula (30) become $\mathbf{x} = (\chi_x, t_x, t_z, \omega_y)$ and $\tilde{\mathbf{x}} = \mathbf{x} + (\Delta\chi_x, \Delta t_x, \Delta t_z, \Delta\omega_y)$. Specializing (30) with respect to (6) leads to

$$\left\|\frac{\partial Z}{\partial\chi_x}(\tilde{\mathbf{x}})\right\| \cdot \|\Delta\chi_x\| + \left\|\frac{\partial Z}{\partial t_x}(\tilde{\mathbf{x}})\right\| \cdot \|\Delta t_x\| + \left\|\frac{\partial Z}{\partial t_z}(\tilde{\mathbf{x}})\right\| \cdot \|\Delta t_z\| + \left\|\frac{\partial Z}{\partial\omega_y}(\tilde{\mathbf{x}})\right\| \cdot \|\Delta\omega_y\|. \quad (35)$$

The term $\frac{\partial Z}{\partial\chi_x}(\tilde{\mathbf{x}})$ evaluates to

$$\frac{t_x - xt_z}{[\chi_x + (x^2+1)\omega_y]^2}. \quad (36)$$

Inspection of (36) shows that its contribution to (35) will be small if three conditions are met: 1) the horizontal disparity χ_x is relatively large; 2) the rotation about the Y-axis, ω_y is relatively large; and 3) the difference of the two relative view translations, t_x and t_z, is relatively small. Intuitively, these observations suggest that stable situations result from those viewing conditions that tend to maximize the difference in the two stereo views. Similarly, the $\frac{\partial Z}{\partial\omega_y}(\tilde{\mathbf{x}})$ term of (35)

evaluates to

$$\frac{t_x(x^2+1)-t_z(x^3+x)}{[\chi_x+(x^2+1)\omega_y]^2}$$

which can be seen to have the same conditions for small magnitude as does (36). Finally, $\frac{\partial Z}{\partial t_x}(\tilde{\mathbf{x}})$ and $\frac{\partial Z}{\partial t_z}(\tilde{\mathbf{x}})$ yield

$$-[\chi_x+(x^2+1)\omega_y]^{-1}$$

and

$$x[\chi_x+(x^2+1)\omega_y]^{-1},$$

respectively. For these last two cases only the conditions that both χ_x and ω_y are relatively large are necessary for stability. On the whole it can be concluded that the magnitude (35) can be kept small (and hence the recovery of r stable) provided that viewing conditions are chosen to make the difference in the two stereo views salient.

The final developments of this section consider the stability of the surface orientation measures as embodied in (22). For these cases the parameters of the generalized error-propagation formula (30) become $\mathbf{x}=(\sigma,\theta,r,t_x)$ and $\tilde{\mathbf{x}}=\mathbf{x}+(\Delta\sigma,\Delta\theta,\Delta r,\Delta t_x)$. (As earlier $(\cos\theta,\sin\theta)$ specifies the direction of the eigenvector $\boldsymbol{\xi}$.) The measure of stability for p can be written

$$\left\|\frac{\partial p}{\partial\sigma}(\tilde{\mathbf{x}})\right\|\cdot\|\Delta\sigma\|+\left\|\frac{\partial p}{\partial\theta}(\tilde{\mathbf{x}})\right\|\cdot\|\Delta\theta\|+\left\|\frac{\partial p}{\partial r}(\tilde{\mathbf{x}})\right\|\cdot\|\Delta r\|+\left\|\frac{\partial p}{\partial t_x}(\tilde{\mathbf{x}})\right\|\cdot\|\Delta t_x\|. \quad (37)$$

The terms of interest in (37) evaluate to

$$\frac{\partial p}{\partial\sigma}(\tilde{\mathbf{x}})=\cos 2\tilde{\theta}\frac{\tilde{r}}{\tilde{t_x}}$$
$$\frac{\partial p}{\partial\theta}(\tilde{\mathbf{x}})=-2\sin 2\tilde{\theta}\frac{\tilde{r}\tilde{\sigma}}{\tilde{t_x}}$$
$$\frac{\partial p}{\partial r}(\tilde{\mathbf{x}})=\cos 2\tilde{\theta}\frac{\tilde{\sigma}}{\tilde{t_x}}$$
$$\frac{\partial p}{\partial t_x}(\tilde{\mathbf{x}})=-\cos 2\tilde{\theta}\frac{\tilde{r}\tilde{\sigma}}{\tilde{t_x}^2}$$

Similarly, the relation for q substituted into (30) leads to consideration of

$$\left\|\frac{\partial q}{\partial\sigma}(\tilde{\mathbf{x}})\right\|\cdot\|\Delta\sigma\|+\left\|\frac{\partial q}{\partial\theta}(\tilde{\mathbf{x}})\right\|\cdot\|\Delta\theta\|+\left\|\frac{\partial q}{\partial r}(\tilde{\mathbf{x}})\right\|\cdot\|\Delta r\|+\left\|\frac{\partial q}{\partial t_x}(\tilde{\mathbf{x}})\right\|\cdot\|\Delta t_x\|. \quad (38)$$

For (38) it is found that

$$\frac{\partial q}{\partial\sigma}(\tilde{\mathbf{x}})=\sin 2\tilde{\theta}\frac{\tilde{r}}{\tilde{t_x}}$$
$$\frac{\partial q}{\partial\theta}(\tilde{\mathbf{x}})=2\cos 2\tilde{\theta}\frac{\tilde{r}\tilde{\sigma}}{\tilde{t_x}}$$
$$\frac{\partial q}{\partial r}(\tilde{\mathbf{x}})=\sin 2\tilde{\theta}\frac{\tilde{\sigma}}{\tilde{t_x}}$$
$$\frac{\partial q}{\partial t_x}(\tilde{\mathbf{x}})=-\sin 2\tilde{\theta}\frac{\tilde{r}\tilde{\sigma}}{\tilde{t_x}^2}$$

Examining the expansions of the terms in (37) and (38) leads to the conclusion that stable solutions are reached when r is relatively small while t_x is relatively large. The requirement that t_x have relatively large magnitude is consistent with earlier results on the importance of keeping the stereoscopic differences salient. The importance of r not being too large reflects the fact that differences in relative surface orientation become less salient as viewing distance increases.

Stepping back from the analysis once again allows for additional positive conclusions. The observations made on the stable recovery of the geometric surface parameters are in accord with the earlier stability results: The key conditions leading to stable recovery of both view and surface geometry center around making the differential viewing information salient; good stereo viewing conditions lead to good solutions.

C. Recapitulation

This section has focused on developing an understanding of how the proposed approach to disparity interpretation can be expected to behave in the face of imperfect data. The discussion began by considering the possibility of degenerate sets of image measurements that would not allow the computations to be defined. It was concluded that for general stereo viewing, degenerate conditions are quite limited and easily diagnosed or avoided. The second set of developments considered the numerical stability of the proposed computations. This analysis resulted in the intuitively pleasing result that stereo viewing conditions that make disparity salient will lead to good stability in the recovery methods. Significantly, these results indicate that an algorithm for recovering scene geometry from stereo disparity can practically monitor its own performance.

IV. Applications

In this section the formal results relating disparity and three-dimensional surface geometry are applied to two particular tasks. The first task is the recovery of three-dimensional surface orientation. The second task is the recovery of three-dimensional surface discontinuities through first-order, i.e., local discontinuities in the geometric surface parameters p, q, and r. These tasks are chosen due to the usefulness of surface orientation and discontinuity information for navigation, object recognition and manipulation.

A. Three-Dimensional Surface Orientation

The algorithm for recovering three-dimensional surface orientation from stereo disparity employs (22) for mapping measurements of orientational disparity to three-dimensional surface orientation. In order to exploit redundant data (i.e., data measurements beyond the minimal number required for a solution) an approach based on histograming is employed. Histograming has been chosen as it makes relatively weak assumptions about the nature of the error component corrupting the data.

There are three main steps to the surface orientation recovery algorithm.

Algorithm for Recovering Surface Orientation

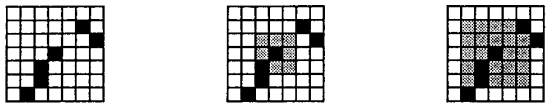


Fig. 5. An algorithm for selecting local regions for input to the surface parameter and histogram computations. Scan by rasters until a line segment is located in the left image; define the line's position by its top left pixel. Then search the eight-connectivity neighborhood pixels for another line segment. If no more segments are found, scan the eight-connected neighbors of the last scanned set, and so on until the desired number of inputs have been scanned. In this example, it takes two iterations of the algorithm to locate the second line segment. Line segments are depicted with black; expanding search regions are depicted as dotted areas.

1) Locally recover values for the three-dimensional surface orientation parameters p and q from an input disparity map.
2) Combine the local values for p and q into separate histograms and smooth.
3) Select the peaks of the p and q histograms as the two parameters specifying surface orientation.

Step 1 of the algorithm decomposes into three subparts: First, locally select three pairs of orientational disparity measurements to serve as input to the p and q recovery. A simple approach based on eight-connectivity serves to define these local groupings of disparity measurements; see Fig. 5. Second, use (20) to map the local orientational disparity measurements into local values for σ and ξ_1. Third, use the equations in (22) to map the recovered values of σ and ξ_1 into local values for surface orientation p and q. Step 2 of the algorithm is a straightforward matter of histogram formation. The selection of the measurements serving as input to a local pair of histograms (one each for p and q) employs the same eight-connectivity approach used to define local inputs to the surface parameter computation. The values in the histogram are then smoothed by forming the serial products of the histogram buckets with the one-dimensional mask [0.5,1.0,0.5]. The smoothing operation helps to reduce discretization errors that are due to the discontinuous nature of the histogram buckets. Step 3 of the algorithm simply selects the peaks of the p and q histograms as the values that specify surface orientation.

The entire algorithm for recovering three-dimensional surface orientation from stereo disparity has been the subject of a software implementation in Zetalisp running on a Symbolics Lisp Machine. An example illustrates application of the algorithm to a natural image stereo pair. The stereo pair is shown in the top half of Fig. 6. The three-dimensional subject matter consists of a textured plane rotated about a vertical axis by approximately 45° with respect to the view direction. Binocular disparity was recovered from this stereo pair in a two step process. First, edge features were extracted using a version of the Canny edge detector [6]. Second, orientational disparity was recovered via a version of the Marr–Poggio–Grimson stereo matcher that defines its matches according to linear segments (cf. earlier versions of this algorithm, Grimson [10]). In applying the surface orientation recovery algorithm to this data, measurements were combined across the image region corresponding to the three-dimensionally oriented plane, subsequent to bringing locally recovered p and q values into a global coordinate system. To facilitate comparison with the stereo pair, the results of the experiment are given in terms of the direction of surface gradient, $\frac{(p,q)}{\|(p,q)\|}$, and the magnitude of surface gradient, $\|(p,q)\|$. The lower left panel of Fig. 6 overlays on the original left image a line corresponding to the recovered direction of surface gradient. The lower right panel of Fig. 6 shows a line slanted by the recovered magnitude of surface gradient. Inspection of the results confirms the adequacy of the proposed algorithm for recovering three-dimensional surface orientation directly from measurements of binocular orientational disparity.

B. Three-Dimensional Surface Discontinuities

The algorithm for recovering three-dimensional surface discontinuities from stereo disparity builds directly on the accumulated results of this paper. In particular, formal relations have been derived for recovering surface orientation p and q and distance, r, from stereo disparity. Further, the relations for recovering surface orientation have been evaluated empirically. Therefore, it should be possible to recover the discontinuities of surfaces through first-order by directly comparing adjacent recovered values for geometric surface parameters, p, q, and r. If adjacent sets of values differ, then a surface discontinuity can be inferred to lie in the region separating the values. In order to take advantage of this idea, there must be a principled approach to the question of what constitutes a "significant" difference between adjacent measures. A statistical test of the difference between samples can be used to answer this question. (Although there might still be an ad hoc element in selecting the level of significance.) Further, a nonparametric method is best for current concerns as it requires minimal commitment to the distribution of the compared parameters and the form of the error component. The nonparametric Kolmogorov–Smirnov method is used here as it can be shown to be both very efficient and powerful (Siegel [29]). The Kolmogorov–Smirnov method is a two-sample test of whether two samples have been drawn from the same source. A large deviation between two sample cumulative distributions is taken as evidence that the samples were drawn from distinct sources. More precisely, let $x_1 < x_2 < \cdots < x_n$ and $y_1 < y_2 < \cdots < y_m$ be the ordered samples from two sources that have continuous cumulative distribution functions $F(z)$ and $G(z)$. Also, let $S_n^x(z) = \frac{k}{n}$ with k the number of samples less than or equal to z for the set x_i. Similarly, let $S_m^y(z) = \frac{l}{m}$ for the set y_i. Then the measure

$$D = \max_z \|S_n^x(z) - S_m^y(z)\| \tag{39}$$

can be used to decide if $F(z) \equiv G(z)$. For small sample size and $n = m$ the probability that $D \leq \frac{h}{n}, h = \max \|k - l\|$, can be derived via a set of recurrence relations. The results of this computation are commonly available in sources such as Siegel [29].

There are three main steps to the surface orientation recovery algorithm.

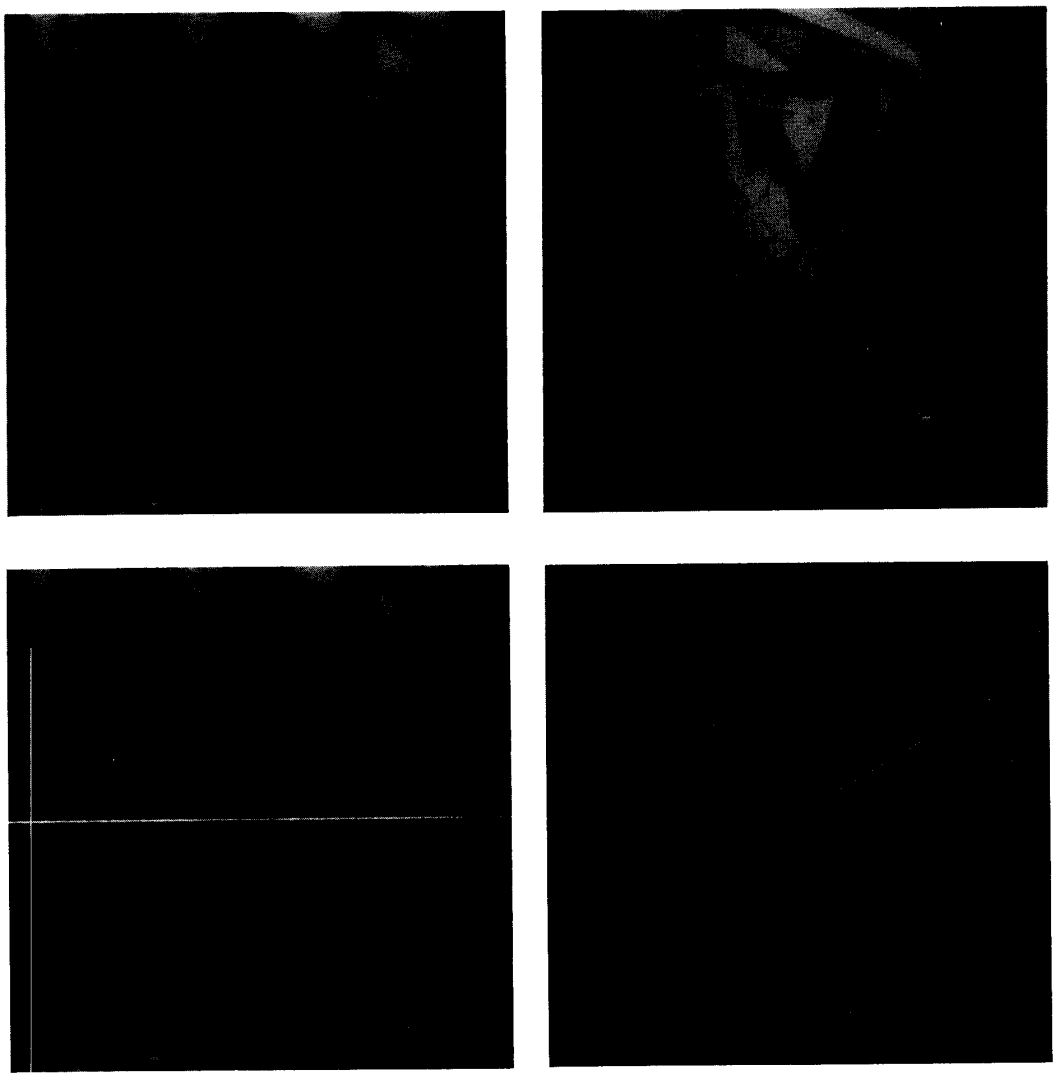


Fig. 6. Results of a computer experiment with the proposed algorithm for recovering surface orientation. The top half of the figure displays a natural image stereo pair. The lower left panel overlays on the original left image a line corresponding to the recovered direction of surface gradient. The lower right panel shows a line slanted by the recovered magnitude of surface gradient.

Algorithm for Recovering Surface Discontinuities

1) Locally recover values for the geometric surface parameters p, q, and r from an input disparity map.
2) Combine the recovered values of p, q, and r into separate local histograms and smooth.
3) Compare adjacent histograms for each surface parameter with the Kolmogorov–Smirnov test. If the value of the test exceeds a specified value then assert a discontinuity in the region between the histograms.

Step 1 of the algorithm recovers values for p and q in same manner described in the algorithm for recovering surface orientation. The value for r is recovered in an analogous fashion while making use of (6) to map horizontal disparity measurements into distance. Similarly, step 2 forms its histograms in the same fashion as described in the algorithm for recovering surface orientation. However, now an additional histogram must be formed for r. Step 3 is concerned with testing the significance of differing p, q, and r between adjacent histogramed regions. This step consists of applying the Kolmogorov–Smirnov test to the neighboring histograms. For each histogram: First, form a cumulative distribution. Second, compute the maximum difference, D, between neighboring cumulative distributions corresponding to each surface parameter. Third, if any value of D exceeds the specified level of significance then assert a discontinuity. For present purposes a discontinuity is asserted to lie in the region between the support areas of the neighboring histograms. Neighboring regions are selected using eight-connectivity.

The entire algorithm for recovering three-dimensional surface discontinuities from stereo disparity has been the sub-

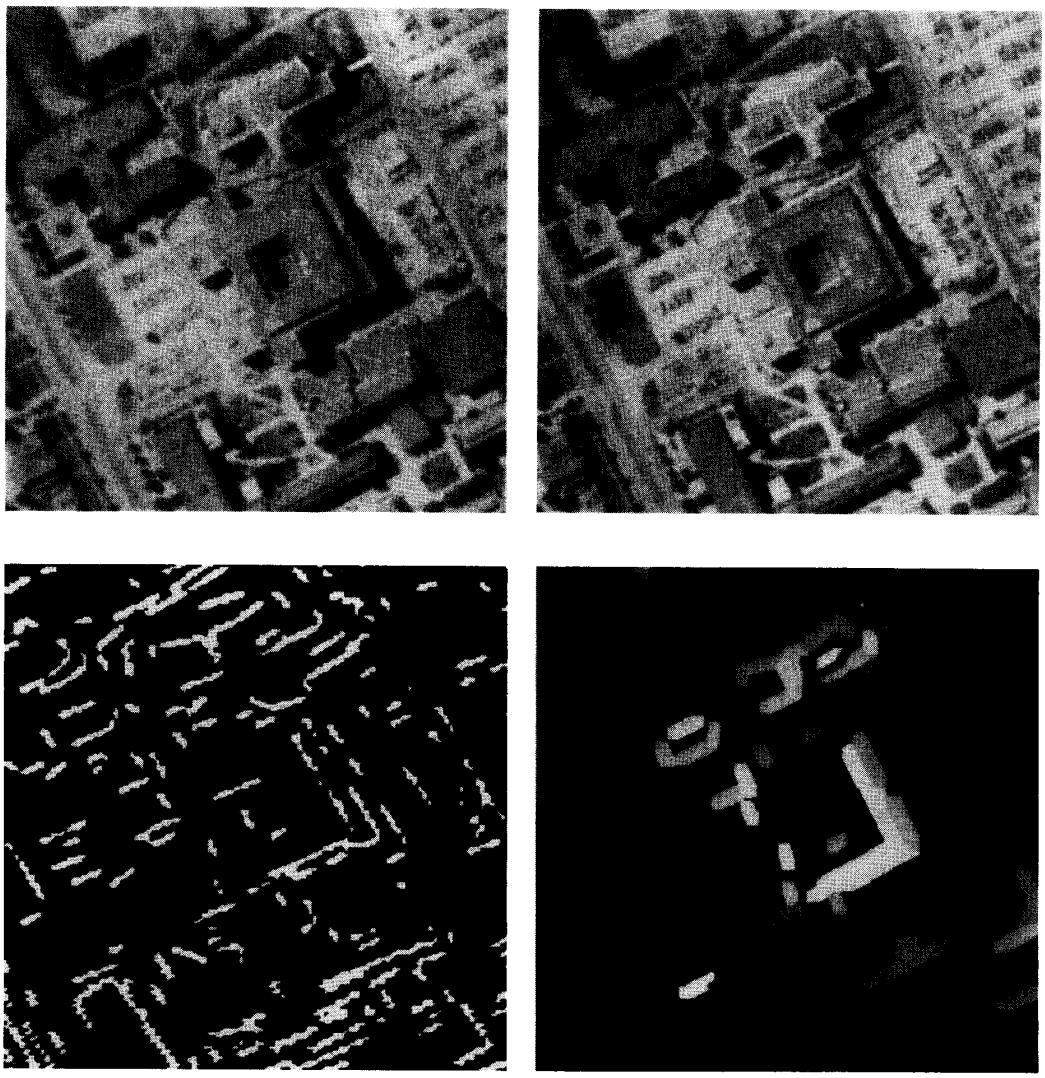


Fig. 7. Results of a computer experiment with the proposed algorithm for recovering the discontinuities of surfaces. The top half of the figure displays a natural image stereo pair. The lower left panel shows the corresponding linear segments from which the horizontal and orientational disparity information was derived. The lower right panel shows the recovered regions of discontinuity. The recovered distance from the viewer is displayed in terms of gray levels with black the furthest and white the closest.

ject of a software implementation in Zetalisp running on a Symbolics Lisp Machine. An example illustrates application of the algorithm to a natural image stereo pair. The stereo pair is shown in the top half of Fig. 7. The three-dimensional subject matter consists of an aerial view of buildings arranged on a ground plane. Binocular disparity was recovered from this stereo pair in a fashion analogous to the surface orientation example.[3] The bottom left panel of the figure shows the linear segments that were used to derive the horizontal and orientational disparity for input to the surface discontinuity detection algorithm. In applying the surface discontinuity recovery algorithm to the disparity measurements, the significance level was set at $D = 0.9$ and histograms were formed over 3 local regions. Further, the intial discontinuity detection was based on local estimates of surface parameters; subsequently parameters along the forward edge of the discontinuity were brought into a global coordinate system allowing for explicit recovery of three-dimensional configurations. Before examining the results of the experiment, it is worth emphasizing that this is a particularly difficult test case for three reasons: 1) There is a significant gradient of disparity across the image. (The buildings are situated on a hill.) Therefore, a simple threshold on raw disparity could not accurately detect discontinuities. 2) The disparity range is rather small. This leads to a poor signal to noise ratio in the input data. 3) The disparity data is quite sparse. The results of the experiment are shown in the lower right panel of Fig. 7. As

[3]The disparity map was provided by E. Grimson of the M.I.T. A.I. Lab.

a convention for displaying the results, the recovered regions of discontinuity are coded in gray levels, with dark gray the furthest and white the closest. Inspection of the results allows for several observations. First, the important discontinuities in distance that are present in the disparity data are recovered by the algorithm. Further, few false alarms are signaled by the algorithm. Second, not all regions of discontinuity are recovered by the algorithm (e.g., certain portions of the outline of the central building). However, reference to the input data in the lower right panel of Fig. 7 shows that there is little data available to the algorithm in these regions. Third, the recovery of relative distance along the discontinuities (as coded by gray level) is in good accord with the corresponding three-dimensional scene geometry.

C. Recapitulation

This section has presented applications of the analysis of stereo disparity that was developed in earlier sections of the paper. In particular, algorithms for recovering three-dimensional surface orientation and discontinuities have been presented. Experiments with these algorithms also were presented. The results of the experiments are positive; the algorithms perform well in recovering the relevant three-dimensional surface geometry in the presented cases.

V. Summary

This paper has presented an analysis of stereo disparity for the recovery of three-dimensional scene geometry. More specifically, a set of relations were derived that explicitly relate the geometry of a stereo disparity map to the geometry of differentially projected surfaces. These relations were used to show how three-dimensional surface geometry through first-order (i.e., relative distance and orientation of surfaces with respect to a viewer) and binocular viewing parameters can be recovered directly from horizontal and orientational disparities. Following these initial developments, a stability analysis of the recovery methods was presented. The results of this analysis show that the recovery methods can be expected to be robust in the face of error perturbed disparity data. Further, the stability analysis indicates how the recovery methods can monitor the reliability of their own behavior. As applications of these formal results, algorithms were presented for recovering three-dimensional surface orientation and surface discontinuities from stereo disparity. Finally, results of applying these algorithms to natural image stereo disparity were presented.

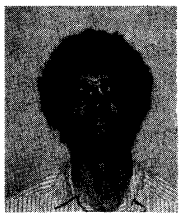
Acknowledgment

E. Grimson, E. Hildreth, W. Richards, and W. Thompson provided comments that improved the quality of research that is presented in this paper.

References

[1] P. Anandan, "A unified perspective on computational techniques for the measurement of visual motion" in *Proc. Int. Conf. Computer Vision*, 1987, pp. 462–472.
[2] S. T. Barnard and M. A. Fischler, "Computational stereo," *Comput. Surveys*, vol. 14, no. 4, pp. 553–572, 1982.
[3] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.
[4] T. E. Boult, "Reproducing kernels for surface interpolation," Columbia Univ., New York, NY, A.I. Memo, 1987.
[5] J. M. Brady, J. Ponce, A. Yuille, and H. Asada, " Describing surfaces," Massachusetts Inst. Technol., Cambridge, A.I. Memo 822, 1985.
[6] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 6, pp. 679–698, 1986.
[7] W. F. Clocksin, "Perception of surface slant and edge labels from optical flow," *Perception*, vol. 9, pp. 253–269, 1980.
[8] G. Dahlquist and Å. Björk, *Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1974.
[9] R. D. Eastman and A. M. Waxman, "Using disparity functionals for stereo correspondence and surface reconstruction," *Comput. Vision, Graphics, Image Processing*, vol. 39, pp. 73–101, 1987.
[10] W. E. L. Grimson, *From Images to Surfaces*. Cambridge, MA: MIT Press, 1981.
[11] W. E. L. Grimson and T. Pavlidis, "Discontinuity detection for visual surface reconstruction," *Comput. Vision, Graphics, Image Processing*, vol. 30, pp. 316–330, 1985.
[12] J. G. Harris, "The coupled depth slope approach to surface reconstruction," in *Proc. Int. Conf. Computer Vision*, 1987, pp. 277–283.
[13] W. Hoff and N. Ahuja, "Extracting surfaces from stereo images," in *Proc. Int. Conf. Computer Vision*, 1987, pp. 284–294.
[14] B. K. P. Horn, "Relative orientation," *Int. J. Computer Vision*, vol. 4, no. 1, pp. 59–78, 1990.
[15] K. Kanatani, "Structure from motion without correspondence: General principle," in *Proc. Int. Joint Conf. Artificial Intelligence*, 1985, pp. 886–888.
[16] C. Koch, J. Marroquin, and A. Yuille, "Analog 'Neural' Networks in Early Vision," Massachusetts Inst. Technol., Cambridge, A.I. Memo 751, 1985.
[17] J. J. Koenderink and A. J. van Doorn, "Geometry of binocular vision and a model for stereopsis," *Biol. Cybern.*, vol. 21, pp. 29–35, 1976.
[18] ——, "Local structure of movement parallax of the plane," *J. Opt. Soc. Amer.*, vol. 66, no. 7, pp. 717–723, 1976.
[19] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers*. New York: McGraw-Hill, 1961.
[20] D. Lee and T. Pavlidis, "One-dimensional regularization with discontinuities," in *Proc. Int. Conf. Computer Vision*, 1987, pp. 572–577.
[21] H. C. Longuet-Higgins and K. Pradzny, "The interpretation of a moving retinal image," *Proc. Roy. Soc. London B*, vol. 208, pp. 385–397, 1980.
[22] *Manual of Photogrammetry*. Washington, DC: Amer. Soc. Photogrammetry (no author or editor named), 1966.
[23] J. L. Marroquin, "Surface reconstruction preserving discontinuities," Massachusetts Inst. Technol., Cambridge, A.I. Memo 792, 1984.
[24] J. E. W. Mayhew and H. C. Longuet-Higgins, "A computational model of binocular depth perception," *Nature*, vol. 297, pp. 376–378, 1982.
[25] G. Medioni and R. Nevatia, "Description of three-dimensional surfaces using curvature properties," in *Proc. DARPA Image Understanding Workshop*, 1984, pp. 219–229.
[26] K. M. Mutch and W. B. Thompson, "Analysis of accretion and deletion at boundaries in dynamic scenes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 7, no. 2, pp. 133–138, 1985.
[27] K. Pradzny, "Egomotion and relative depth map from optical flow," *Biol. Cybern.* vol. 36, pp. 87–102, 1980.
[28] B. Schunk, "The motion constraint equation for optical flow," in *Proc. Int. Joint Conf. Pattern Recognition*, 1986, pp. 20–22.
[29] S. Siegel, *Nonparametric statistics*. New York: McGraw-Hill, 1956.
[30] T. Smitley and R. Bajcsy, "Stereo processing of aeriel urban images," in *Proc. Int. Joint Conf. Pattern Recognition*, 1984, pp. 405–409.
[31] K. A. Stevens and A. Brookes, "Depth reconstruction in stereopsis," in *Proc. Int. Conf. Computer Vision*, 1987, pp. 549–603.
[32] M. Subbarao, "Interpretation of image flow," *Int. J. Comput. Vision*, vol. 2, no. 1, pp. 77–96, 1988.
[33] D. Terzopoulos, "Regularization of inverse problems involving discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 4, pp. 413–424, 1986.
[34] ——, "Integrating visual information from multiple sources," in *From Pixels to Predicates*, S. Pentland, Ed. Norwood, NJ: Ablex, 1986.

[35] W. B. Thompson, K. M. Mutch, and V. A. Berzins, "Edge detection in optical flow fields," in *Proc. Amer. Assoc. Artificial Intelligence*, 1982, pp. 26–29.
[36] ——, "Dynamic occlusion analysis in optical flow fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, no. 4, pp. 374–383, 1985.
[37] C. Truesdell and R. A. Toupin, "The classical field theories," in *Handbuch der Physik*, S. Flügge, Ed. Berlin: Springer-Verlag, 1960.
[38] A. Waxman and S. Ullman, "Surface structure and three-dimensional motion from image flow kinematics," *Int. J. Robotics Res.*, vol. 4, no. 3, pp. 72–94, 1985.
[39] D. Weinshall, "Qualitative depth and shape from stereo in agreement with psychophysical evidence," Massachusetts Inst. Technol., Cambridge, A.I. Memo 1007, 1987.
[40] R. P. Wildes, "On interpreting stereo disparity," Massachusetts Inst. Technol., Cambridge, A.I. Lab. Tech. Rep. 1112, 1989.
[41] Y. Yeshurun and E. L. Schwartz, "Cepstral filtering on a columnar image architecture: A fast algorithm for binocular stereo segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 7, pp. 759–767, 1989.



Richard P. Wildes (S'88–M'88) received the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 1989.

From 1984 through 1988, he was a research assistant in the M.I.T. Artificial Intelligence Laboratory and the Department of Brain and Cognitive Sciences. During that time, he was a National Science Foundation Graduate Fellow. In 1988, he joined the Department of Computer Science at the State University of New York at Buffalo as an Assistant Professor. During 1990, he joined the S.R.I. David Sarnoff Research Center, Princeton, NJ, where he is a member of the technical staff. His main areas of research interest are machine and biological perception (especially vision), robotics, and artificial intelligence.