# Direct Estimation of Local Surface Shape in a Fixating Binocular Vision System*

Jonas Gårding and Tony Lindeberg

Computational Vision and Active Perception Laboratory (CVAP)
Department of Numerical Analysis and Computing Science
Royal Institute of Technology (KTH), S-100 44 Stockholm, Sweden
Email: jonasg@bion.kth.se, tony@bion.kth.se

**Abstract.** This paper addresses the problem of computing cues to the three-dimensional structure of surfaces in the world directly from the local structure of the brightness pattern of a binocular image pair. The geometric information content of the gradient of binocular disparity is analyzed for the general case of a fixating system with symmetric or asymmetric vergence, and with either known or unknown viewing geometry. A computationally inexpensive technique which exploits this analysis is proposed. This technique allows a local estimate of surface orientation to be computed directly from the local statistics of the left and right image brightness gradients, without iterations or search. The viability of the approach is demonstrated with experimental results for both synthetic and natural gray-level images.

## 1 Introduction

Binocular disparities, i.e., the slight differences between the views of the world captured by the left and the right eye, can convey important information about the three-dimensional structure of objects and surfaces in the scene. Traditionally, binocular stereopsis has often been associated with recovery of three-dimensional *depth.* Here, however, we shall be concerned with estimation of *surface orientation*, i.e., the rate of change of depth. Many computational models of stereopsis are based on sparse but salient features such as edges or corners (see e.g. (Pollard *et al.* 1985)). This approach is often quite successful, but has the drawback that it only produces sparse depth estimates. If higher-order properties are needed, such as local surface orientation or curvature, they could in principle be estimated by first applying an additional stage that interpolates the surface between the data points to obtain a dense depth map (see e.g. Blake and Zisserman 1987), and then differentiating this representation.

An alternative approach, which we shall pursue here, is to derive higher-order surface properties directly from the properties of corresponding image patches,

* This work was partially performed under the ESPRIT-BRA project INSIGHT, and with support from the Swedish National Board for Industrial and Technical Development, NUTEK. Part of the work was performed during a visit by the first author to the AI Vision Research Unit, Sheffield University.

Lecture Notes in Computer Science, Vol. 800
Jan-Olof Eklundh (Ed.)
Computer Vision – ECCV '94
© Springer-Verlag Berlin Heidelberg 1994

without using depth as an intermediate representation. This can be achieved either by first computing a dense disparity map and then estimating derivatives of the disparity field, or by directly using differences in local image properties, e.g. the local statistics of the orientation or curvature of contours.

In both cases, the estimation of surface orientation can be formulated in terms of modelling the local transformation from the right eye's view of a small surface patch to the left eye's view of the same patch by an *affine* transformation, rather than a simple displacement. Analogously, surface curvature can be estimated from the second-order properties of the local left-to-right transformation. The local affine transformation gives rise to *orientation disparity* as well as *spatial frequency disparity*, and several computational models based more or less directly on these cues have been described in the literature (Blakemore 1970; Koenderink and van Doorn 1976; Tyler and Sutter 1979; Rogers and Cagenello 1989; Wildes 1991; Jones and Malik 1992).

The present work builds on most of these theories, and extends them in several ways. In the first part of the paper, we analyze the geometric structure of the problem. We first treat the case in which the orientation of the cameras is known, and then generalize to the case of unknown camera orientation which allows the surface shape to be recovered up to the group of relief transformations. In the second part of the paper, we propose a direct and inexpensive computational technique which exploits the geometric analysis. This technique allows a local estimate of surface orientation to be computed directly from the local statistics of the left and right image brightness gradients, without iterations or search.

## 2 Viewing Geometry and Binocular Disparity

**Viewing Geometry.** A representation of the binocular viewing geometry is shown in Figure 1. We represent visual space with respect to a virtual cyclopean eye, constructed such that the cyclopean visual axis (the $Z$ axis) bisects the left and right visual axes. The $X$ and $Z$ axes as well as the centres of the eyes lie in a common plane, called the fixation plane.

We define left and right coordinate systems $(X_l, Y_l, Z_l)$ and $(X_r, Y_r, Z_r)$ such that the origin of each system is at the center of projection, the $Z_l$, $Z_r$ and $Z$ axes intersect at the fixation point $p$ with cyclopean coordinates $(0, 0, R)$, and the $X_l$, $X_r$ and $X$ axes are contained in the fixation plane. Normalized cyclopean image coordinates are defined by $x = X/Z$, $y = Y/Z$; left and right image coordinates are defined analogously. These coordinates are related to the pixel coordinates through the intrinsic camera parameters, which are assumed to be known.

This representation of the viewing geometry does not require $p$ to be the actual fixation point of the viewing system, nor indeed that the eyes fixate any point at all, since a rotation of either eye around the optical centre does not affect the information content of the image. However, to simplify the presentation we shall continue to refer to $p$ as the fixation point.

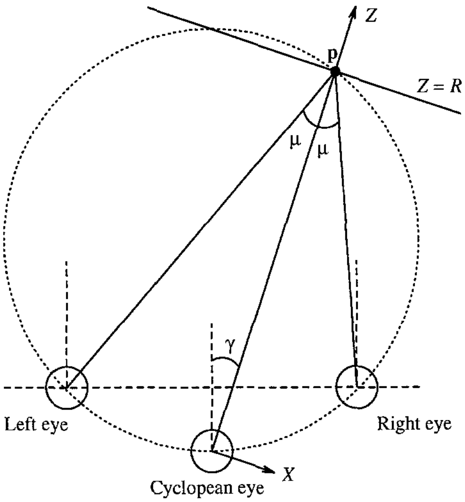

**Fig. 1.** Representation of the binocular viewing geometry. The plane of the drawing is the fixation plane. The primary direction (indicated by dashed lines) is defined as the direction in the fixation plane that is perpendicular to the interocular baseline. The dotted circle through the fixation point and the eyes indicates a part of the point horopter, i.e., the locus of points that yield zero horizontal and vertical disparity.

**Vergence and Version.** Let $\varphi_l$ and $\varphi_r$ be the angles between the primary (straight-ahead) direction and the left and right visual axes respectively. The *vergence* angle $\mu$ and the *version* (or *gaze*) angle $\gamma$ are then defined by

$$\mu = \frac{1}{2}(\varphi_l - \varphi_r), \qquad \gamma = \frac{1}{2}(\varphi_l + \varphi_r).$$

As a consequence of this definition, the angle between the cyclopean visual axis and the primary direction is equal to $\gamma$ (see Figure 1).

It is sometimes convenient to specify the cyclopean fixation distance $R$ instead of the vergence angle $\mu$. The relation between these parameters is

$$R = \frac{I\ \cos\gamma}{\sin 2\mu}, \tag{1}$$

where $I$ is the interocular distance.

**Camera Orientation.** The orientation of a camera with respect to a reference system in the "head" has three degrees of freedom; two for the orientation of the optical axis, and one for the *torsion*, i.e., the angle of rotation around the visual axis. For human vision, Donder's law states that the eyes do not use the third degree of freedom; the amount of torsion is fully determined by the direction of the visual axis for each eye. This reduces the number of degrees of freedom to four for the whole binocular system. One additional constraint is supplied by the assumption that the eyes fixate some point $p$; the total number of degrees of freedom for the orientation of the binocular system is thus three, i.e., the extrinsic geometry of the binocular system is fully determined by the coordinates of the fixation point.

A more convenient way of specifying the extrinsic binocular geometry is by using the angles of vergence and gaze defined previously, supplemented with the angle $\epsilon$ of *elevation* of the fixation plane with respect to a reference plane containing the interocular baseline. This representation, which is equivalent to the Helmholtz fixation model (Helmholtz 1910; Carpenter 1988), has the advantage that the vertical axes of the left and right coordinate systems remain perpendicular to the fixation plane for all fixation points; as a consequence, any computed entity which is defined relative to the fixation plane is independent of $\epsilon$.

**Binocular Disparity.** The retinal disparity of a point in the scene is defined as the difference in retinal position of the left and right projections of the point. Consequently, the retinal disparity of the fixation point is zero by definition. We define horizontal and vertical retinal disparity $(h, v)$ by

$$h = x_r - x_l, \quad v = y_r - y_l,$$

where $(x_l, y_l)$ and $(x_r, y_r)$ are the normalized left and right image coordinates corresponding to the same point in the scene.

If the fixation point $p$ lies on a smooth surface $Z(X, Y)$, a differentiable mapping $M$ is induced from points in the left image to points in the right image in some neighbourhood of the images of $p$. A Taylor expansion to first order in $(x_r, y_r)$ can then be expressed as

$$\begin{pmatrix} x_r \\ y_r \end{pmatrix} = \begin{pmatrix} 1 + h_x & h_y \\ v_x & 1 + v_y \end{pmatrix} \begin{pmatrix} x_l \\ y_l \end{pmatrix} . \tag{2}$$

In the following we shall denote the matrix in (2) by $M_*$ and refer to it as the *derivative map.* The components $(h_x, h_y; v_x, v_y)$ constitute the *disparity gradient.*

## 3 The Disparity Gradient

The disparity gradient depends on the viewing geometry and the local surface orientation. At the fixation point, the precise relation is given by

**Proposition 1 (Disparity gradient).** *Let $M_*$ be the derivative map from the left image to the right image. The disparity gradient is $M_* - I$, where $I$ is the unit matrix, and at the fixation point*

$$M_* = \begin{pmatrix} 1 + h_x & h_y \\ v_x & 1 + v_y \end{pmatrix} = \frac{\cos(\gamma - \mu)}{\cos(\gamma + \mu)} \begin{pmatrix} \dfrac{\cos\mu + P\sin\mu}{\cos\mu - P\sin\mu} & \dfrac{2Q\cos\mu\sin\mu}{\cos\mu - P\sin\mu} \\ 0 & 1 \end{pmatrix} , \tag{3}$$

*where* $P = \frac{\partial Z}{\partial X}$, $Q = \frac{\partial Z}{\partial Y}$.

See (Gårding and Lindeberg 1994b) for a derivation.

The size of the region where $M_*$ provides a reasonably accurate approximation of the disparity field depends on the shape of the surface; for planar surfaces it is in fact valid over quite large visual angles.

### 3.1 The Information Content of the Disparity Gradient

What do the non-vanishing components $(h_x, h_y, v_y)$ of the disparity gradient at the fixation point tell us about the local scene structure and the viewing geometry? First, note that the disparity gradient (3) depends on four parameters; two for the viewing geometry $(\mu, \gamma)$ and two for the surface orientation $(P, Q)$. It is thus impossible to recover both the viewing geometry and the local surface orientation from a single measurement of the disparity gradient. However, if the viewing geometry is known, the surface orientation can be estimated, and vice versa. Moreover, the surface orientation is independent of the gaze angle $\gamma$, since $\gamma$ only affects the overall scale factor in $M_*$ according to (3). Formally, denote the components of $M_*$ by $m_{ij}$, and define the normalized horizontal components as

$$\hat{m}_{11} = m_{11}/m_{22}, \quad \hat{m}_{12} = m_{11}/m_{22}.$$

By comparison with (3) we obtain after some algebraic manipulations

$$P = \frac{(\hat{m}_{11} - 1)\ \cos\mu}{(\hat{m}_{11} + 1)\ \sin\mu}, \qquad Q = \frac{\hat{m}_{12}}{(\hat{m}_{11} + 1)\ \sin\mu}. \tag{4}$$

Consequently, to estimate the surface orientation it suffices to estimate $M_*$ up to an arbitrary scale factor, and there is no need to know the angle $\gamma$ of asymmetric gaze.

### 3.2 Unknown Viewing Geometry and the Relief Ambiguity

An important line of research in computational vision concerns the recovery of three-dimensional structure under "weak calibration" conditions, in which the epipolar geometry is known but the intrinsic camera parameters as well as the extrinsic camera orientation remain unknown. Typically, this allows the scene structure to be recovered up to an arbitrary projective or affine transformation (Koenderink and van Doorn 1991; Faugeras 1992; Robert and Faugeras 1993).

In a fixating binocular system, however, the extrinsic camera orientation is quite constrained; as pointed out in Section 2, it has essentially only three degrees of freedom. Moreover, these angles vary continuously as the system changes its fixation in the visual field, so unlike the remaining parameters of the system they could not be even approximately determined by a preliminary calibration stage. It is therefore of interest to study the case in which only these *dynamic* parameters (i.e., the angles of vergence, gaze and elevation) of a binocular vision system are unknown.[2]

[2] In fact, the subsequent analysis would also allow *some* of the intrinsic camera parameters to be unknown, but we shall not pursue this possibility further here.

In fact, two of the three dynamic degrees of freedom have already been eliminated; the elevation angle by assuming Helmholtz fixation, and the gaze angle by normalizing the disparity gradient. Hence, we only need to analyze the influence of the vergence $\mu$.

It is convenient to introduce the *small baseline approximation*,[3] which applied to some expression $f$ is defined to be the term(s) up to first order in a Taylor expansion of $f$ with respect to $I/R$. Rearranging terms in (4) and then applying this approximation, we obtain

$$2\frac{(\hat{m}_{11}-1)}{(\hat{m}_{11}+1)} = 2P\tan\mu \approx (I\cos\gamma)\,\frac{P}{R}, \tag{5}$$

$$2\frac{\hat{m}_{12}}{(\hat{m}_{11}+1)} = 2Q\sin\mu \approx (I\cos\gamma)\,\frac{Q}{R}, \tag{6}$$

where (1) has been used to expand $\tan\mu$ and $\sin\mu$ to first order with respect to $I/R$. The right-hand sides of these expressions have an interesting geometric interpretation in terms of *nearness*, i.e., inverse depth. At the origin of the cyclopean system we have

$$\frac{\partial}{\partial x}\frac{1}{Z} = -\frac{1}{R}\frac{\partial Z}{\partial X} = -\frac{P}{R}, \qquad \frac{\partial}{\partial y}\frac{1}{Z} = -\frac{1}{R}\frac{\partial Z}{\partial Y} = -\frac{Q}{R}.$$

Hence, at the fixation point the gradient of nearness can be computed up to the scale factor $I\cos\gamma$ without any knowledge about the viewing geometry. Suppose that this scaled gradient has been computed in a region around the fixation point.[4] By integration we can then recover the function

$$\rho(x,y) = I\cos\gamma\left(\frac{1}{Z(x,y)} - \frac{1}{R}\right), \tag{7}$$

where $I$, $\gamma$ and $R$ can be considered as unknown constants. We shall refer to $\rho$ as *scaled relative nearness.*

Clearly, knowledge of $\rho$ determines the scene structure up to a two-fold ambiguity corresponding to the unknown parameters $A = I\cos\gamma$ and $B = 1/R$. This ambiguity has a clean mathematical structure which allows a simple geometric interpretation. Consider an arbitrary member of this family of scene configurations, obtained from some arbitrarily chosen values $(A', B')$, and denote the true values by $(A, B)$. It is then easily verified that the position $(X', Y', Z')$ of any point in this scene configuration is related to the position $(X, Y, Z)$ of the corresponding point in the true scene configuration by

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \frac{1}{a+bZ}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{8}$$

[3] This approximation can in fact be justified even for quite small viewing distances; see (Gårding and Lindeberg 1994b).

[4] A straightforward application of the method in the whole visual field yields the gradient of inverse cyclopean distance rather than inverse depth. A suitable coordinate transformation converts one of these representations to the other.

where $a = A/A'$ and $b = B' - (A/A')B$. We shall refer to this as a *relief transformation*; it is an instance of what is sometimes referred to as the *bas-relief ambiguity*. Examples of the effect of the relief transformation (8) are shown in Figure 2.

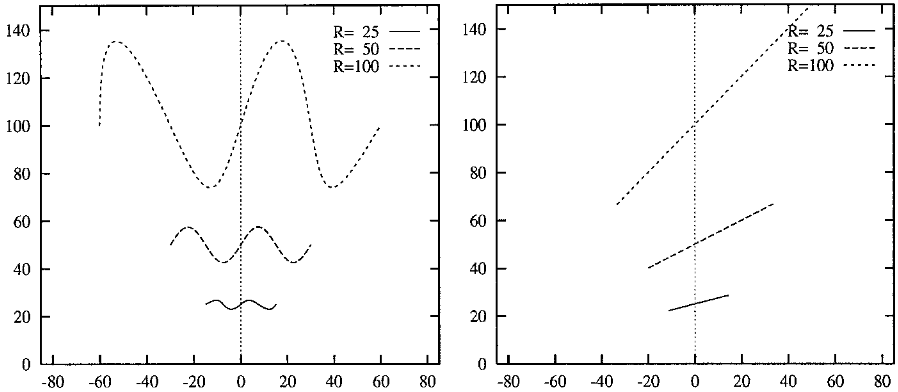

**Fig. 2.** Relief transformations. The diagrams show horizontal cross-sections of a group of surfaces related by the relief transformation (8). In these examples the parameter $B = 1/R$ is varied whereas $A = I\cos\gamma$ is held fixed. The position of the cyclopean eye is at $(0, 0)$. Note that planes are mapped to planes, and that the depth ordering is preserved.

Koenderink and van Doorn (1976, 1991) have pointed out that many aspects of perceived visual shape are invariant against relief transformations. This fact was noted already by Helmholtz (1910), and artists have long exploited it. To develop these remarks formally, we first note that (8) is a linear transformation in projective three-space $\mathbb{P}^3$, which means that it preserves coplanarity and collinearity. Moreover, if the scene is thought of as consisting of a stack of "depth planes" of constant $Z$, the transformation (8) preserves the *ordering* of these planes. Consequently, shape judgements such as planar–curved or convex–concave can be performed without resolving the relief ambiguity. A useful and intuitively appealing way of understanding (8) is as an *equivalence class* of three-dimensional shapes; this is well-defined since it defines a transformation group.

The idea of representing the structure of the three-dimensional environment up to a relief transformation has also been applied to unify theories of binocular stereopsis in human vision (Gårding et al 1994).

**Cyclotorsion and Disparity Deformation.** The preceding analysis is similar but not equivalent to the "disparity deformation" model proposed by Koenderink and van Doorn (1976). Both methods are based on the small baseline approximation, but the deformation model also allows arbitrary cyclotorsion around the line of sight of each camera. To obtain this invariance, however, it is necessary to estimate the full structure of $M_*$ up to scale (i.e., three parameters), unlike the method proposed above which only uses the normalized horizontal components of $M_*$ (i.e., two parameters). This difference will turn out to be of great practical importance for the computational technique that will be described next.

# 4 Direct Estimation of the Disparity Gradient

The preceding analysis has shown how to interpret the disparity gradient under conditions of known or unknown dynamic viewing geometry. These results can be applied to dense disparity fields computed by any stereo matching method, but in the following we shall use them for *direct* estimation of the disparity gradient, without first establishing a large number of point correspondences. We shall only assume the ability to fixate, i.e., to establish correspondence for a single point. Moreover, this correspondence will be allowed to be approximate.

Basically, the technique by which this will be achieved is to compute a certain descriptor of the structure of the local brightness pattern in the left and right image patches, and then to use the difference between these descriptors to compute the required parameters of the local affine transformation between the patches. This method is an adaptation of a computational framework for estimation of shape-from-texture proposed in (Lindeberg and Gårding 1993; Gårding and Lindeberg 1994a), which will be briefly reviewed below.

This approach differs from those based on orientational disparity (Koenderink and van Doorn 1976; Wildes 1991) in that it does not require a preprocessing step in which contours are extracted from the image; rather, it is based directly on the outputs of simple local operators (more precisely, first-order Gaussian derivatives). In this respect it is therefore similar to the filter-based approach proposed by Jones and Malik (1992), but whereas in that approach the parameters of the local left-to-right affine transformation are estimated by exhaustive search in the space of permissible transformations, we derive a closed-form expression for the transformation parameters directly in terms of the operator outputs.

## 4.1 The Windowed Second Moment Descriptor

A simple image descriptor that allows estimation of linear spatial distortion can be defined as follows. Let $L\colon \mathbb{R}^2 \to \mathbb{R}$ denote the image brightness and let $\nabla L = (L_x, L_y)^T$ be its gradient. Given a symmetric and normalized Gaussian window function $w$, *the windowed second moment matrix* $\mu_L$ can be defined as

$$\mu_L(q) = \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{pmatrix} = E_q \begin{pmatrix} L_x^2 & L_xL_y \\ L_xL_y & L_y^2 \end{pmatrix} = E_q((\nabla L)(\nabla L)^T), \qquad (9)$$

where $E_q$ is an averaging operator describing the effect of integration with the window function $w$ centered at $q$. Different versions of it have been used by several authors; see e.g. (Lindeberg and Gårding 1993) for a review.

*Transformation property.* Let $B$ be an invertible linear transformation of the image domain and define a transformed intensity pattern $R\colon \mathbb{R}^2 \to \mathbb{R}$ by $L(\xi) = R(B\xi)$. Then, it can be shown that $\mu_L(q)$ transforms according to

$$\mu_L(q) = B^T \, \mu_R(p) \, B, \tag{10}$$

where $\mu_R(p)$ is the second moment matrix of $R$ at $p = Bq$ computed with respect to the backprojection of the window function $w$.

*Directional statistics.* The trace of a second moment descriptor $\mu_L$ is equal to the average squared gradient magnitude. The remaining two degrees of freedom of the descriptor contain *directional* information, which can be represented by

$$\tilde{C} = (\mu_{11} - \mu_{22})/\text{trace}\ \mu, \qquad \tilde{S} = 2\,\mu_{12}/\text{trace}\ \mu. \tag{11}$$

It is easily verified that $(\tilde{C}, \tilde{S})^T$ is invariant with respect to uniform scaling of either brightness $L$ or the spatial coordinates $(x, y)$. The computational technique described below uses only these directional components of $\mu_L$.

### 4.2 Estimating Surface Orientation

Let $\mu_L$ and $\mu_R$ denote the windowed second moment matrices computed at the left and right images of the fixation point. If the linearized mapping from the left to the right image is denoted by $M_*$, then from (10)

$$\mu_L = M_*^T \, \mu_R \, M_*. \tag{12}$$

If $\mu_L$ and $\mu_R$ are known, then (12) provides three equations for the four parameters of the linear transformation $M_*$; it can be shown that the general solution to (12) is

$$M_* = \mu_R^{-1/2} W^T \mu_L^{1/2} \tag{13}$$

where $W$ is an arbitrary orthogonal matrix, and the notation $\mu^{1/2}$ indicates the unique positive definite symmetric solution to the equation $X^2 = \mu$. Here, however, the viewing geometry provides the additional constraint $m_{21} = v_x = 0$ (assuming no cyclotorsion), so in this case it is in fact possible to recover $M_*$ completely from $\mu_L$ and $\mu_R$ (excluding degenerate cases).

To recover the surface orientation, however, only the normalized horizontal components $\hat{m}_{11} = m_{11}/m_{22}$ and $\hat{m}_{12} = m_{11}/m_{22}$ are needed (see Section 3). These components can be computed from the difference in the *directional* structure of $\mu_L$ and $\mu_R$, while ignoring any difference in magnitude. As pointed out earlier, this procedure has the additional advantage that there is no need to know the angle of asymmetric gaze.

Expressing $M_*$ in terms of $(\hat{m}_{11}, \hat{m}_{12}, m_{22})$ and using $m_{21} = 0$, (12) can be rewritten

$$\mu_L = m_{22}^2 \begin{pmatrix} \hat{m}_{11}^2 \mu_{11}^R & \hat{m}_{11}(\hat{m}_{12}\mu_{11}^R + \mu_{12}^R) \\ \hat{m}_{11}(\hat{m}_{12}\mu_{11}^R + \mu_{12}^R) & \hat{m}_{12}^2 \mu_{11}^R + 2\hat{m}_{12}\mu_{12}^R + \mu_{22}^R \end{pmatrix}, \tag{14}$$

where $\mu_{ij}^R$ denotes the components of $\mu_R$. Substituting the directional components $(\tilde{C}, \tilde{S})^T$ defined by (11) into (14), we obtain after some algebraic manipulation

$$\hat{m}_{11} = \frac{1 + \tilde{C}_L}{1 + \tilde{C}_R} \frac{\tilde{F}_R}{\tilde{F}_L}, \qquad \hat{m}_{12} = \frac{\tilde{S}_L \tilde{F}_R - \tilde{S}_R \tilde{F}_L}{(1 + \tilde{C}_R) \tilde{F}_L}, \tag{15}$$

where

$$\tilde{F}_L = \sqrt{1 - \tilde{C}_L^2 - \tilde{S}_L^2}, \qquad \tilde{F}_R = \sqrt{1 - \tilde{C}_R^2 - \tilde{S}_R^2}.$$

## 5 Experimental Results

Figure 3 shows the ellipse representation of the windowed second moment matrix computed at the fixation point and four neighbouring points superimposed on a bright copy of a synthetic stereo pair (arranged for cross-eyed fusion). The images are perspective views of a sinusoidal pattern, and contain 5% additive Gaussian noise. The visual angle across the diagonal of each image is 32°, and the vergence angle is $\mu = 10°$. The orientation of the surface is $(\hat{P} = 1, \hat{Q} = \sqrt{2})$.

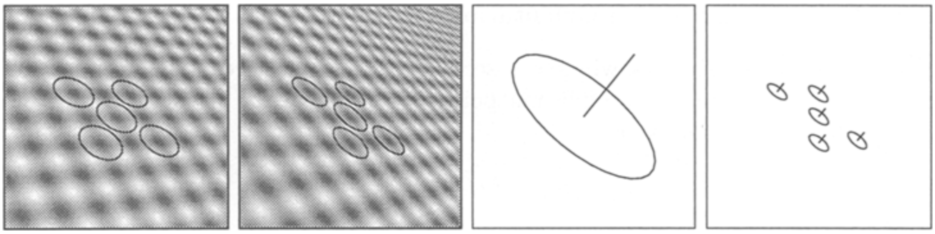

**Fig. 3.** Local surface orientation estimated from the gradient of horizontal disparity in a synthetic stereo pair with 5% noise. The columns show from left to right; (a-b) Bright copies of the right and left images with the computed texture descriptor superimposed. (c) reference surface orientation, (d) estimated surface orientation at five manually matched points. (c) and (d) are shown with respect to the left eye's view.

At the fixation point, the estimated normalized horizontal disparity gradient was $(\hat{m}_{11} = 1.405, \hat{m}_{12} = 0.577)$, and from (4) we then obtain the estimated surface orientation $(P = 0.96, Q = 1.38)$. The error in the estimate, expressed as the angle between the estimated and true surface normals, is only 0.9°. The results obtained at the remaining four points were very similar, as can be seen from the graphical representation shown to the right in Figure 3.

Figure 4 shows the results obtained by applying the same procedure to a real stereo pair. A number of point pairs were matched manually, and $\mu_L$ and $\mu_R$ were then computed at each of these points. Together the estimates clearly indicate the curved shape of the object, although a few of the individual estimates contain significant errors.

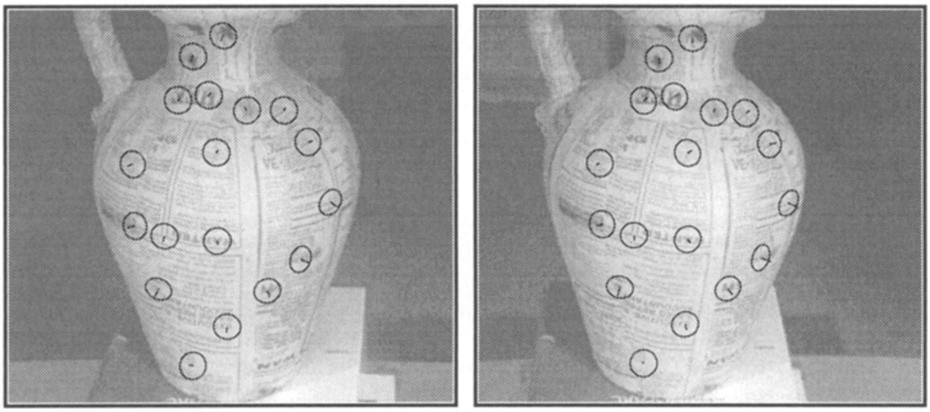

**Fig. 4.** Local surface orientation estimated from the gradient of horizontal disparity in a real stereo image of a curved object (arranged for cross-eyed fusion). The estimates are shown with respect to both the right and the left views.

## 6 Conclusions

We have analyzed the geometric information content of the gradient of binocular disparity, both for the cases of known and unknown dynamic viewing geometry. If the vergence angle is known, the disparity gradient can be used to recover local surface orientation independently of the gaze angle. If the vergence angle is unknown, the disparity gradient determines the three-dimensional structure of the scene up to a relief transformation, which preserves projective properties as well as depth ordering. As an application of the geometric analysis, we presented a direct and inexpensive computational technique which allows a local estimate of surface orientation to be computed directly from the local statistics of the left and right image brightness gradients, without iterations or search.

The direct method described in Section 4 uses a very limited amount of information to estimate the affine transformation between two image patches. The performance in terms of accuracy can therefore not be expected to match that which can be obtained by more elaborate and computationally intensive methods; the value of the direct approach lies in the fact that it makes hypotheses about local orientation available with a few simple low-level operations and a limited computational effort. As needed, each hypothesis could then be verified and improved by a separate mechanism. A complementary way of improving the accuracy is described in (Lindeberg and Gårding 1994).

## References

A. Blake and A. Zisserman, *Visual Reconstruction.* MIT Press, Cambridge, Mass., 1987.

C. Blakemore, "A new kind of stereoscopic vision", *Vision Research*, vol. 10, pp. 1181–1200, 1970.

R.H.S. Carpenter, *Movements of the Eyes.* Pion Limited, London, second ed., 1988.

O. Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig?", in *Proc. 2nd European Conf. on Computer Vision* (G. Sandini, ed.), vol. 588 of *Lecture Notes in Computer Science*, pp. 563–578, Springer-Verlag, May 1992.

J. Gårding and T. Lindeberg, "Direct computation of shape cues by multi-scale retinotopic processing", *Int. J. of Computer Vision*, 1994a. (To appear).

J. Gårding and T. Lindeberg, "Direct estimation of local surface shape in a fixating binocular vision system", Tech. Rep. TRITA-NA-P9408, Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, 1994b.

J. Gårding, J. Porrill, J.E.W. Mayhew, and J.P. Frisby, "Binocular stereopsis, vertical disparity and relief transformations", *Vision Research*, 1994. (To appear).

H.L.F. von Helmholtz, *Treatise on Physiological Optics*, vol. 3. (trans. J.P.C Southall, Dover, New York 1962), 1910.

D.G. Jones and J. Malik, "Determining three-dimensional shape from orientation and spatial frequency disparities", in *Proc. 2nd European Conf. on Computer Vision* (G. Sandini, ed.), vol. 588 of *Lecture Notes in Computer Science*, pp. 661–669, Springer-Verlag, May 1992.

J.J. Koenderink and A.J. van Doorn, "Geometry of binocular vision and a model for stereopsis", *Biological Cybernetics*, vol. 21, pp. 29–35, 1976.

J.J. Koenderink and A.J. van Doorn, "Affine structure from motion", *J. of the Optical Society of America A*, vol. 8, pp. 377–385, 1991.

T. Lindeberg and J. Gårding, "Shape from texture from a multi-scale perspective", in *Proc. 4th Int. Conf. on Computer Vision*, (Berlin, Germany), pp. 683–691, May 1993.

T. Lindeberg and J. Gårding, "Shape-adapted smoothing in estimation of 3-D depth cues from affine distortions of local 2-D brightness structure", in *Proc. 3rd European Conf. on Computer Vision*, (Stockholm, Sweden), May 1994.

S.B. Pollard, J.E.W. Mayhew, and J.P. Frisby, "PMF: A stereo correspondence algorithm using a disparity gradient limit", *Perception*, vol. 14, pp. 449–470, 1985.

L. Robert and O. Faugeras, "Relative 3D positioning and 3D convex hull computation from a weakly calibrated stereo pair", in *Proc. 4th Int. Conf. on Computer Vision*, (Berlin, Germany), pp. 540–544, May 1993.

B.J. Rogers and R. Cagenello, "Orientation and curvature disparities in the perception of three-dimensional surfaces", *Investigative Opthalmology and Visual Science*, vol. 30, p. 262, 1989.

C.W. Tyler and E.E. Sutter, "Depth from spatial frequency difference: An old kind of stereopsis?", *Vision Research*, vol. 19, pp. 859–865, 1979.

R.P. Wildes, "Direct recovery of three-dimensional scene geometry from binocular stereo disparity", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 13, no. 8, pp. 761–774, 1991.