

Stereopsis, vertical disparity and relief transformations*

J Gårding¹, J Porrill², J E W Mayhew², J P Frisby²

Computational Vision and Active
(1) Perception Laboratory (CVAP) (2) AI Vision Research Unit
Royal Institute of Technology (KTH) Sheffield University, UK
Stockholm, Sweden

November 1993; revised June 1994

Keywords: binocular vision, stereopsis, vertical disparity, 3-D shape perception, depth

Abstract

The pattern of retinal binocular disparities acquired by a fixating visual system depends on both the depth structure of the scene and the viewing geometry. This paper treats the problem of interpreting the disparity pattern in terms of scene structure without relying on estimates of fixation position from eye movement control and proprioception mechanisms. We propose a sequential decomposition of this interpretation process into *disparity correction*, which is used to compute three-dimensional structure up to a relief transformation, and *disparity normalization*, which is used to resolve the relief ambiguity to obtain metric structure. We point out that the disparity normalization stage can often be omitted, since relief transformations preserve important properties such as depth ordering and coplanarity. Based on this framework we analyze three previously proposed computational models of disparity processing; the MLH model (Mayhew and Longuet-Higgins 1982), the deformation model (Koenderink and van Doorn 1976) and the polar angle disparity model (Weinshall 1990). We show how these models are related, and argue that none of them can account satisfactorily for available psychophysical data. We therefore propose an alternative model, *regional disparity correction*. Using this model we derive predictions for a number of experiments based on vertical disparity manipulations, and compare them to available experimental data. The paper is concluded with a summary and a discussion of the possible architectures and mechanisms underlying stereopsis in the human visual system.

*This work was supported by the ESPRIT-BRA project INSIGHT. The first author gratefully acknowledges the support from the Swedish Institute and the Foundation Blanceflor Boncompagni-Ludovisi, née Bildt. We thank David Buckley and Stephen Hippisley-Cox for valuable comments and discussions. Finally, we thank the anonymous reviewers for many useful suggestions which significantly improved the presentation of this work.

Contents

1	Introduction	1
2	Viewing geometry and binocular disparity	3
2.1	Viewing geometry	3
	Vergence and version	3
	Torsion	4
2.2	Binocular disparity	4
	Disparity, viewing geometry and depth	5
2.3	Epipolar geometry	7
3	Disparity correction and 3-D shape	8
3.1	Nearness and relief transformations	8
3.2	Relation to weak calibration models	11
4	Computational models of vertical disparity processing	11
4.1	The Mayhew and Longuet-Higgins (MLH) model	12
4.2	The disparity deformation (def) model	12
4.3	The polar angle disparity (PAD) model	13
	Relation to the def model	16
4.4	The need for an improved model	16
4.5	The regional disparity correction (RDC) model	17
	Relation to the polar angle disparity model	19
5	Vertical disparity and psychophysics	19
5.1	Some general observations	19
5.2	Unilateral vertical magnification: the induced effect	20
5.3	Vertical disparity scaling	21
	Perceived shape of the fronto-parallel plane	22
	Shape judgements and the horizontal/vertical anisotropy	23
5.4	Vertical disparity and psychophysics: Summary	25
6	Summary and discussion	26
6.1	Summary	26
6.2	Discussion	26
	Architecture of stereopsis	26
	The scale of vertical disparity processing	28
A	The disparity equations	29
B	The PAD model: A psychophysical experiment	29

1 Introduction

Computation of three-dimensional structure from binocular disparities can be viewed as a two-stage process. In the first stage, correspondence is established between the projections of features onto the retinae of the two eyes, and in the second stage the disparities, i.e., the differences between left and right retinal positions, are interpreted in terms of the three-dimensional structure of objects and surfaces in the scene. Any computational model of this process must take into account the fact that both stages are highly dependent on the viewing geometry, i.e., the orientation of the eyes relative to the head. In the matching stage, the viewing geometry manifests itself in the form of the *epipolar constraint*, i.e., the fact that any point on one retina corresponding to a given point on the other must lie on the intersection of the retina with the plane defined by the given point and the left and right projection centers. This constraint allows the search for corresponding points to be carried out in a one-dimensional rather than a two-dimensional space, thus reducing both the computational complexity and the probability of an incorrect match. In the interpretation stage, a single disparity vector taken in isolation is highly ambiguous, but knowledge of the viewing geometry allows the three-dimensional location of a given feature to be determined by intersecting the rays from the left and right retinae.

In some binocular machine vision systems, the viewing geometry is fixed (e.g. with approximately parallel cameras) and can be determined once and for all by a calibration procedure. In human vision, however, or indeed any fixating visual system, the viewing geometry changes continually as the gaze is shifted from point to point in the visual field. In principle, this situation can be approached in two different ways; either a mechanism must be provided which continuously makes the state of the viewing geometry available to the binocular system, or invariant representations that fully or partially side-step the need for calibration of the viewing geometry must be found. For each approach a number of different techniques are possible, and any combination of these may be used as they are not mutually exclusive.

The viewing geometry could in principle be recovered from extraretinal sources, using either in-flow or out-flow signals from the oculomotor and/or accommodation systems. The viability of this approach has been questioned on the ground that judgements of depth from oculomotor/accommodation information alone are notoriously poor (Foley 1980, 1985; review in Collett, Schwartz and Sobel 1991). Alternatively, the viewing geometry can be recovered from purely visual information, using the mutual image positions of a number of matched image features to solve for the rotation and translation of one eye relative to the other. This is often referred to in the photogrammetry and computer vision literature as the “relative orientation” problem, and its mathematical structure is fairly well understood (see e.g. Horn (1990) for a review).

For normal binocular vision the relative orientation problem need not be solved in its full generality since the kinematics of fixating eye movements is quite constrained. These constraints lead to a natural decomposition of the disparity field into a horizontal component, which carries most of the depth information, and a vertical component, which mainly reflects the viewing geometry. Under these circumstances the problem of compensating the horizontal disparity field for the influence of the viewing geometry in order to obtain the metric structure of the scene is traditionally referred to as *disparity scaling*. We shall avoid this term, however, since this process involves much more than a simple multiplicative scaling of the individual disparity values. A computational model of metric reconstruction

from disparities was proposed by Mayhew (1982) and Mayhew and Longuet-Higgins (1982). In this model the viewing parameters needed to solve horizontal disparity for metric scene structure are estimated from the global structure of the vertical disparity field.

A different way of interpreting the local disparity pattern without knowledge of the viewing geometry was proposed by Koenderink and van Doorn (1976). They showed that the gradient of inverse distance to the surface can be computed from the deformation (or *def*) component of the local linear structure of the pattern of horizontal and vertical disparities, and that this component is (to a first approximation) independent of the viewing geometry. An approach in a similar spirit was taken by Weinshall (1990), who proposed a method based on the *polar angle* component of disparity for obtaining a qualitative depth ordering of points in the scene. More recently, the ability of the polar angle disparity (PAD) model to predict the slope of a planar surface has been investigated by Liu, Stevenson and Schor (1994).

Psychophysical research regarding the influence of vertical disparity on stereopsis in the human visual system has a long history beginning with Helmholtz (1910). Many of the computational models that have been proposed are inspired by the work of Ogle (1950) on the induced effect. In recent years, there has been a spur of activity in this area; see e.g. (Petrov 1980; Stenton, Frisby and Mayhew 1984; Gillam, Chambers and Lawergren 1988; Cumming, Johnston and Parker 1991; Sobel and Collett 1991; Rogers 1992; Rogers and Bradshaw 1993).

The interpretation of these results is complicated by the fact that it is not clear what should constitute the “end result” of binocular stereopsis. The Mayhew and Longuet-Higgins (1982) model shows how full metric depth constancy can be achieved. In contrast, the *def* and polar angle models only determine surface structure up to the so-called *relief ambiguity*, which means that only the relative *nearness*, i.e., inverse distance, of points in the scene is determined. To compute the full metric structure of the scene it is still necessary to know the viewing parameters (in particular the fixation distance d). Moreover, these models leave the epipolar geometry undetermined, which significantly complicates the correspondence problem. Interestingly, however, a large body of psychophysical evidence (e.g. Collett, Schwartz and Sobel 1991; Johnston 1991) suggests that human performance in tasks involving estimation of metric structure from binocular disparities is remarkably poor, especially in comparison with the exquisite precision attained in stereoacuity tasks which test the ability of making relative disparity comparisons. Nevertheless, there must exist *some* type of compensation for the variable viewing geometry in human binocular vision. For example, a planar surface normal to the cyclopean line of sight is usually perceived as such for a wide range of fixation distances and angles of asymmetric gaze, despite the fact that the pattern of horizontal disparities varies significantly within this range of stimuli.

To clarify these issues, we propose a general sequential decomposition of metric reconstruction from disparities into *disparity correction*, which is the process of computing shape up to a relief transformation, and *disparity normalization*, which is the process of resolving the relief ambiguity to compute metric depth. Disparity correction can be described as a purely additive mechanism that compensates disparity for retinal eccentricity but leaves the fixation-dependent scale factor relating disparities to depth unspecified. It turns out that many visual tasks can be achieved on the basis of disparity correction alone, which means that the disparity normalization step may often be omitted by the visual system.

The structure of the paper is the following. In Section 2 we introduce and define a number of fundamental concepts related to viewing geometry, binocular disparity and relief

transformations. Section 3 analyzes the geometry of disparity correction and the structure of the resulting shape ambiguities. In Section 4 we analyze a number of previously proposed computational models of vertical disparity processing and show how they are related. We argue that none of them can account satisfactorily for available psychophysical data, and propose an alternative model, *regional disparity correction*. In Section 5 we use this model to derive predictions for a number of experiments based on vertical disparity manipulations, and compare these predictions to available experimental data. Finally, Section 6 concludes the paper with a summary and a discussion.

2 Viewing geometry and binocular disparity

2.1 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.

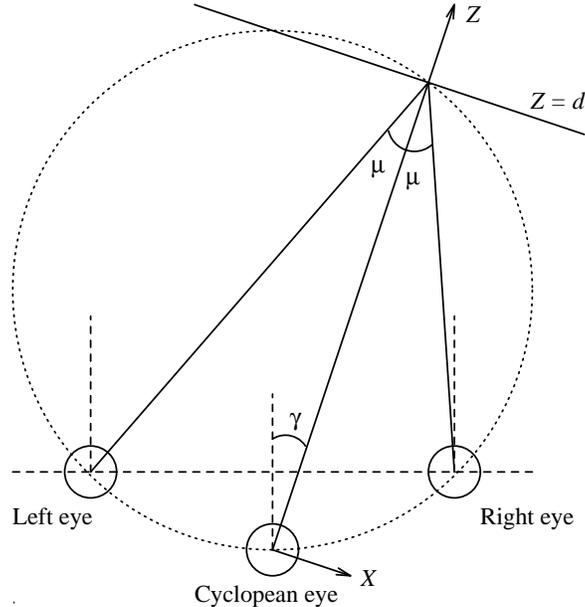


Figure 1: Representation of the binocular viewing geometry. The plane of the drawing is referred to as the fixation plane. The primary direction (indicated by dashed lines) is defined as the direction in the fixation plane which is perpendicular to the interocular baseline. The Vieth-Müller circle (dotted)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

Vergence and version

Let φ_l and φ_r be the angles between the primary (straight-ahead) direction and the left and right visual axes respectively. The *vergence* angle μ and the *version* (or *gaze*) angle γ

are then defined by

$$\begin{aligned}\mu &= \frac{1}{2}(\varphi_l - \varphi_r), \\ \gamma &= \frac{1}{2}(\varphi_l + \varphi_r).\end{aligned}$$

Note in Figure 1 that γ is the angle between the cyclopean visual axis and the primary direction. It is also worth pointing out that this model is slightly different from another commonly used model where the cyclopean eye is placed midway on the line connecting the left and right eyes. The advantage of the cyclopean model in Figure 1 is that the angular relationship between the cyclopean eye and the left and right eyes is completely symmetric; the disadvantage is that the cyclopean eye moves slightly back and forth as a function of the vergence angle. However, except for extremely near fixation positions, the difference between this model and the median model is for all practical purposes negligible.

It is sometimes convenient to specify the cyclopean fixation distance d instead of the vergence angle μ . The relation between these parameters is

$$d = \frac{I \cos \gamma}{\sin 2\mu}, \quad (1)$$

where I is the interocular distance.

Torsion

The representation in Figure 1 does not specify the *torsion* of each eye, i.e., the angle of rotation around its visual axis. For human vision, Donders' law states that the eyes do not use these extra two degrees of freedom; the amount of torsion is fully determined by the direction of the visual axis for each eye. We shall adopt the arbitrary but convenient convention of representing torsion by the angle between the horizontal axis of the eye and the fixation plane. A consequence of this definition is that torsion would be identically zero for all fixation points if the eyes moved according to the Helmholtz model, i.e., by first elevating both eyes to the new fixation plane and then shifting the gaze within this plane to the fixation point. A more realistic model of human eye movements is Listing's law, which states that each eye rotates around an axis perpendicular to the fixation direction and the primary direction; this model predicts non-zero torsion angles for fixation points off the primary plane. (See Helmholtz (1910) or Carpenter (1988) for a more thorough discussion of models of eye movements).

2.2 Binocular disparity

The input signal to stereopsis is the pattern of simultaneous stimulation of pairs of points on the left and right retinae. To formalize this observation, we define the *retinal disparity* of a point in the scene 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.

In order to obtain a numerical representation of disparity we must choose a way of parameterizing retinal position, or equivalently, visual direction relative to each eye. This can be done in a number of different but essentially equivalent ways, e.g. by spherical or projective coordinates. Clearly, this choice is only a matter of convenience which has

nothing to do with the physical shape of the retina. In the following we shall represent a visual direction by standard projective coordinates, which can be thought of as the (x, y) coordinates of the intersection between the visual ray and a planar surface at unit distance from the center of projection. In the cyclopean system we have $x = X/Z$, $y = Y/Z$; left and right image coordinates (x_l, y_l) and (x_r, y_r) are defined analogously. When these coordinates refer to the projections of the same physical point in the scene, the horizontal and vertical retinal disparities of the point are defined by

$$h = x_r - x_l, \quad (2)$$

$$v = y_r - y_l. \quad (3)$$

Another common representation of retinal position is the spherical coordinates $\tan^{-1} x$ and $\tan^{-1}(y/\sqrt{x^2 + 1})$. We shall refer to the associated disparities as *angular* horizontal and vertical disparities. This parameterization has the advantage that it leads to disparity invariances with respect to certain eye movements. In particular, any fixation shift within the same fixation plane preserves the angular vertical disparity of any point, as well as the difference in angular horizontal disparity of any pair of points.

Regardless of the parameterization of visual directions, it is easily seen that no new information can be gained by rotating either eye around its optical centre. In some situations it can therefore be useful to define visual directions and disparity relative to the head rather than to each eye. The difference between the left and right head-centred visual directions to a point in the scene is usually referred to as the *optic array disparity* of the point; this quantity is by definition independent of eye movements and fixation. Some caution is required, however, because unlike retinal disparity, optic array disparity is not something that the visual system can measure directly from its sensory input. Rather, it must be *computed* from retinal disparity by supplying information about eye position and then solving for the absolute orientation of the visual rays. But if the visual system is capable of doing this, it could just as well solve for the exact three-dimensional position of each point directly, without using any disparity representation at all. We therefore hold the view that although optic array disparity can be useful as a geometric construction tool, any *computational* theory of stereopsis should be based on retinal disparity. In the following the term “disparity” will be used as synonymous to “retinal disparity”.

Disparity, viewing geometry and depth

The use of a planar coordinate system has the advantage that disparity fields can easily be visualized. Figure 2 shows examples of how the viewing geometry can affect the disparity field corresponding to a planar surface normal to the cyclopean visual direction. Clearly, if the perceived surface remains invariant, there must exist some mechanism that compensates the disparity field for the varying viewing geometry.

In general, the equations expressing the relationship between horizontal/vertical disparity, viewing geometry and three-dimensional depth are quite complex and not very enlightening. To simplify the algebraic analysis, it is common to make the “small baseline” approximation: assuming that the interocular distance I is small compared to the fixation distance d , only the first-order term in a Taylor expansion with respect to I/d is retained. Let $(X, Y, d + \delta)$ be the cyclopean coordinates of a point in the scene. Expanding h and v

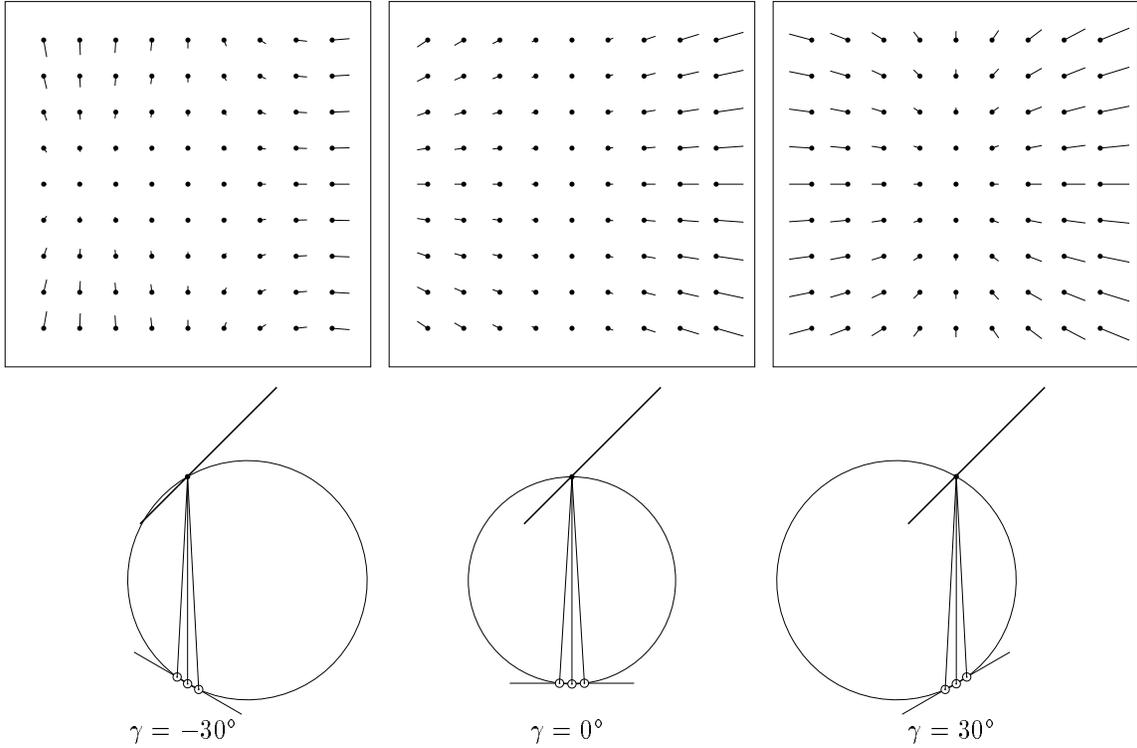


Figure 2: Top: Disparity fields for the same slanted planar surface viewed with different angles of gaze, i.e., by maintaining the fixation point while turning the head. Each vector represents the left (dot) and right (tip) image positions of some point on the surface. **Bottom:** A cross-section of the viewing geometry used to generate the disparity fields. The drawing shows the left, right and cyclopean eyes, as well as the Vieth-Müller circle. The sloping line at the top indicates the planar surface. The fixation distance is 50cm, the interocular distance is 6cm, and the plane is slanted 45° around a vertical axis relative to the cyclopean visual axis. The visual angle subtended by the diagonal of each disparity field is 70.5° .

with respect to I/d in the case of no cyclotorsion, we obtain

$$h = \frac{I \cos \gamma}{d} \left(\frac{\delta}{d + \delta} + \frac{d \tan \gamma}{d + \delta} x + x^2 \right) + \mathcal{O} \left(\left(\frac{I}{d} \right)^3 \right), \quad (4)$$

$$v = \frac{I \cos \gamma}{d} \left(\frac{d \tan \gamma}{d + \delta} y + xy \right) + \mathcal{O} \left(\left(\frac{I}{d} \right)^3 \right), \quad (5)$$

where (x, y) are cyclopean image coordinates (see the Appendix for a derivation). The first thing to note is that the second-order terms vanish; the small baseline approximation is therefore more accurate than one might have guessed. An example of the effects of this approximation on depth reconstruction at a relatively close viewing distance ($d = 25\text{cm}$) is shown in Figure 3. It can be seen e.g. that with a horizontal field of view subtending 23° , the deviation of the approximately reconstructed surface from the true surface is less than 0.1%. Such errors are clearly negligible in comparison to the effects of the various disparity manipulations with which we shall be concerned later on.

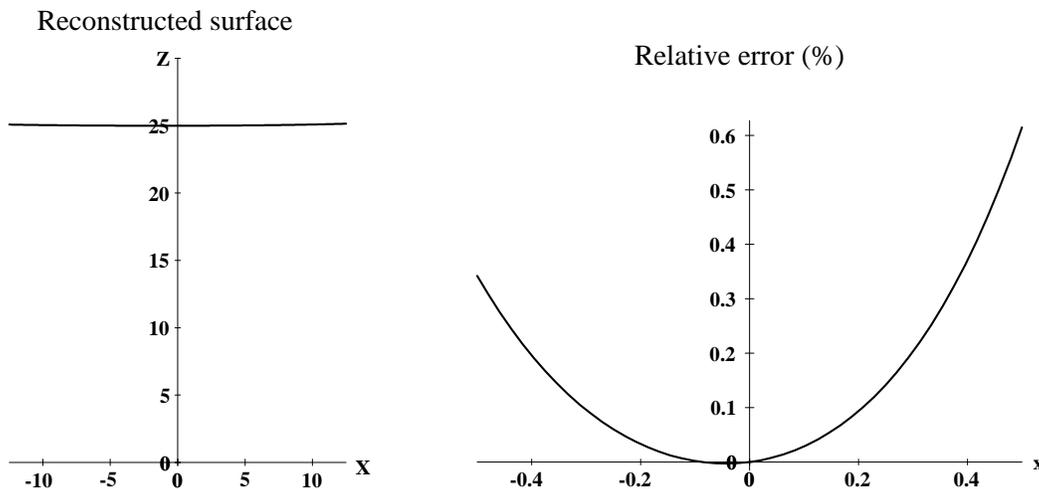


Figure 3: The result of reconstructing the plane $Z = 25\text{cm}$ from horizontal disparities by solving the small baseline approximation (4) for δ . The horizontal visual angle subtended by the surface is 53° , the fixation distance is 25cm , the interocular distance is 6cm and the gaze angle is 5° . **Left:** Intersection of the reconstructed surface with the fixation plane $Y = 0$. The cyclopean eye is at the origin. **Right:** Relative error $\epsilon = 100(\hat{Z} - Z)/Z$, where \hat{Z} is the reconstructed depth, as a function of the cyclopean image coordinate x .

2.3 Epipolar geometry

A point P_r in the right image can correspond to a given point P_l in the left image only if the ray from the right projection center O_r through P_r intersects the ray from the left projection center O_l through P_l . Hence, the possible correspondences to a given point P_l lie along the intersection of the right image and the plane defined by P_l , O_l and O_r . This is known as the *epipolar constraint*, and it simplifies the search for corresponding points considerably. However, as shown in Figure 4 the position and orientation of the epipolar lines depend on the viewing geometry, so in order to exploit this constraint the viewing geometry must in principle be known.

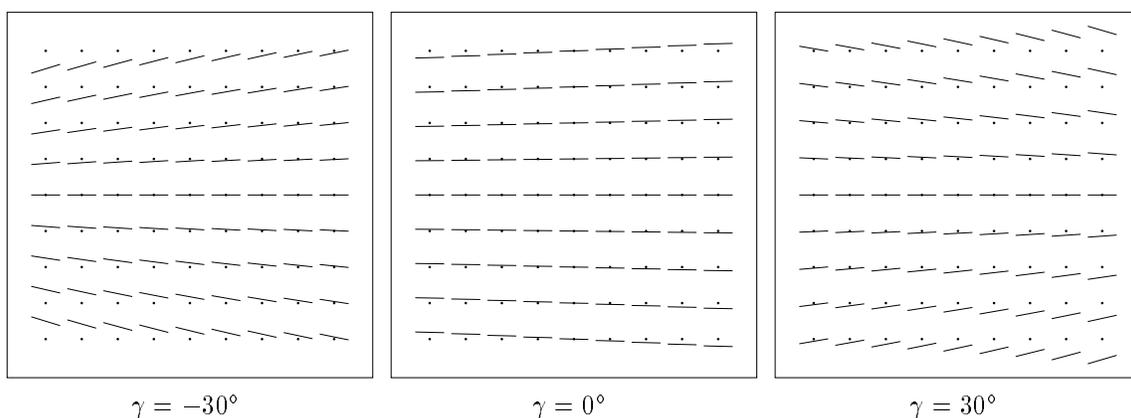


Figure 4: Epipolar geometry with different angles of gaze, when $d = 50\text{cm}$ and $I = 6\text{cm}$. The dots represent left image positions, and for each point the line segment indicates the position and orientation of the epipolar line, i.e., the physically possible corresponding points in the right image.

3 Disparity correction and 3-D shape

It is not a priori clear what the end result of stereopsis should be. Perhaps the first idea that springs to mind is that it should provide a full reconstruction of the metric structure of the scene. There are at least two problems with this idea, however. First, several psychophysical investigations (e.g. Collett, Schwartz and Sobel 1991; Johnston 1991) have found that human performance in tasks involving estimation of metric structure from binocular disparities is remarkably poor, even in the presence of a richly structured disparity field. Second, it turns out that the visual tasks that actually require a full metric reconstruction of the scene are fairly uncommon. Some tasks such as hand-eye coordination only require alignments in depth, rather than absolute depth judgements, and can therefore be driven directly by the raw disparities. Other slightly more elaborate tasks which most humans can perform fairly well, e.g. judgements of planarity and relative depth, can operate on non-metric representations based on relative *nearness*, i.e., the difference in inverse distance. There exists a very natural relationship between such relative nearness representations and stereopsis, which we shall now explore in more detail.

3.1 Nearness and relief transformations

It was pointed out in the introduction that the process of recovering metric depth from binocular disparities can be logically divided into a first step, *disparity correction*, in which an affine function of nearness (i.e., inverse distance) is computed, and a second step, *disparity normalization*, in which a relief transformation based on the estimated viewing geometry is applied to the nearness representation to yield metric depth. We shall now take a closer look at some general properties of the first of these two steps.

The nearness of an arbitrary point (X, Y, Z) is defined as

$$\lambda = 1/Z = 1/(d + \delta).$$

The nearness of the fixation point is thus

$$\lambda_0 = 1/d.$$

The disparity equations (4) and (5) can then be expressed as

$$h = I \cos \gamma (\lambda_0 - \lambda) + I \sin \gamma \lambda x + I \cos \gamma \lambda_0 x^2, \quad (6)$$

$$v = I \sin \gamma \lambda y + I \cos \gamma \lambda_0 xy. \quad (7)$$

We now define *scaled relative nearness* as

$$\rho = I \cos \gamma (\lambda_0 - \lambda), \quad (8)$$

which is the first term in the horizontal disparity equation (6). Note that $\rho = h$ on the vertical meridian $x = 0$. We define *disparity correction* as the process of recovering ρ from h and v in the whole visual field. As can be seen from (6), this process can be described as *adding* a quantity that depends on the viewing parameters (λ_0, γ) and horizontal eccentricity x to the horizontal disparity h ; in particular, it does not involve any multiplicative scaling of disparity by the estimated fixation distance.

The introduction of ρ may at this point appear somewhat arbitrary, so it is worth pointing out that this representation is motivated both by the fact that (as will be shown

in Section 4) it can be computed in several natural ways without explicitly estimating the viewing parameters, and by the fact that it contains sufficient information for a wide range of perceptual tasks. To elaborate the second of these points, we begin by noting that scaled relative nearness (8) is an instance of the more general definition

$$\rho = A(B - \lambda),$$

where A and B are arbitrary and unknown positive constants. Knowledge of ρ alone obviously does not allow full recovery of three-dimensional point positions, but it significantly restricts the possible configurations of points in the scene. We define *disparity normalization* as the process of choosing some particular values of the viewing parameters $A = I \cos \gamma$ and $B = \lambda_0$ and then solving (8) for Z . This of course determines the full three-dimensional position (X, Y, Z) of the point; X and Y are given by the intersection of the ray through the cyclopean image position (x, y) of the point with the depth plane Z , i.e., $X = xZ$ and $Y = yZ$.

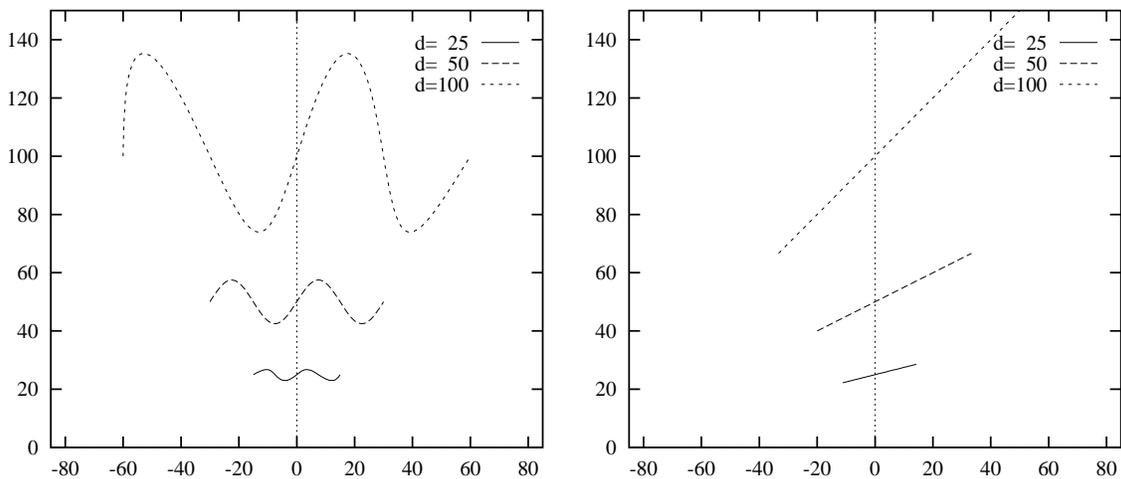


Figure 5: *Relief transformations.* The diagrams show horizontal cross-sections of a group of surfaces related by the relief transformation (9). In these examples the parameter $B = 1/d$ is varied whereas $A = I \cos \gamma$ is held fixed. The position of the cyclopean eye is at $(0, 0)$. Note that planes are mapped to planes, and that the depth ordering remains invariant.

Clearly, knowledge of ρ determines the scene structure up to a two-fold ambiguity corresponding to the unknown parameters $A = I \cos \gamma$ and $B = 1/d$. 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 the transformation

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \frac{1}{a + bZ} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad (9)$$

where $a = A/A'$ and $b = B' - (A/A')B$. We shall refer to (9) as a *relief transformation*. Examples of its geometric effect shown in Figure 5.

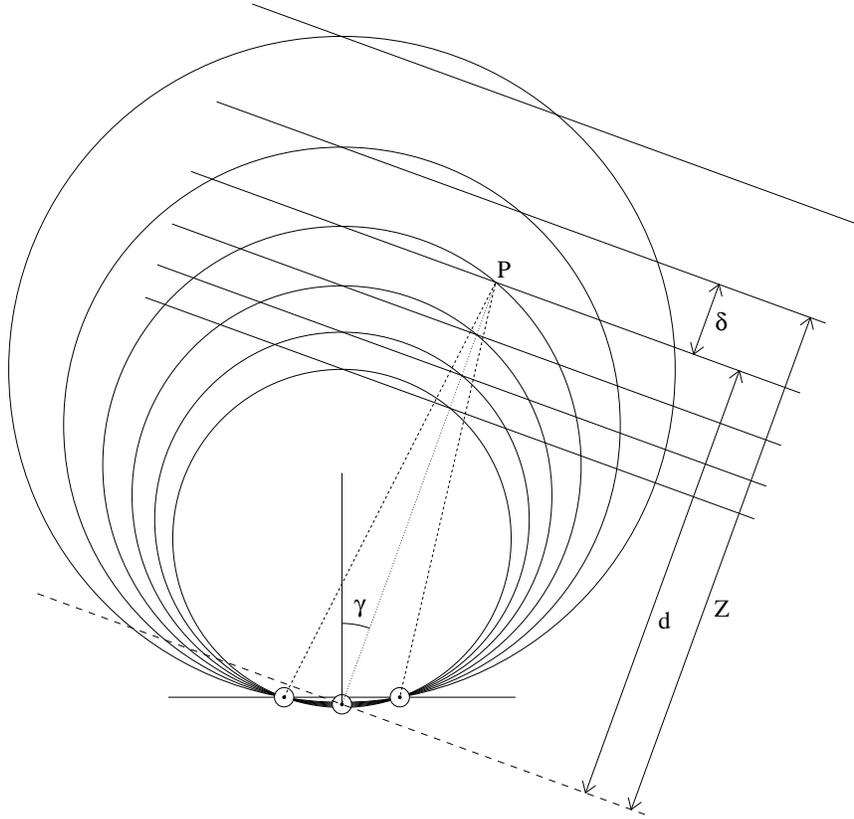


Figure 6: *Disparity correction is the process of computing scaled relative nearness ρ from horizontal disparity h . The Vieth-Müller circles are level curves of angular horizontal disparity, and the straight lines are level curves of ρ . The relation between horizontal disparity and surface structure is complicated by three factors. First, the angle between the cyclopean optical axis and the V-M circles introduces a gaze-dependent horizontal gradient of horizontal disparity (reflected by the x term in (6)); second, the curvature of the V-M circles gives rise to a distance-dependent second derivative of horizontal disparity (reflected by the x^2 term in (6)); and third, the variable spacing between the V-M circles means that the relative depth corresponding to a given amount of horizontal disparity depends on the viewing distance. Disparity correction eliminates the first two of these three factors by straightening the V-M circles into parallel lines, while leaving the interline spacing undetermined.*

Relief transformations were discussed quite extensively by Helmholtz (1910), who pointed out that they preserve many important aspects of visual form, dimensions and shading. The empirical rules governing constructions of relief have been known to artists even longer (Breysig 1798; cited in Helmholtz 1910). More recently, relief transformations¹ have been discussed and analyzed by Koenderink and van Doorn (1976,1986,1991).

To describe the precise mathematical and geometric meaning of the relief transformation (9), let us first note that it is a linear transformation in projective three-space \mathbb{P}^3 . One consequence of this is that it maps planes to planes and straight lines to straight lines. Moreover, if the scene is thought of as consisting of a stack of “depth planes” of

¹The definition (9) corresponds in principle to the relief transformation discussed by Helmholtz (1910), but alternative definitions have also appeared in the literature; for example, in connection with the disparity deformation model, Koenderink and van Doorn (1976) discussed a slightly different kind of relief transformation which is related to concentric spherical shells rather than depth planes.

constant Z , the transformation (9) preserves the *ordering* of these planes. Consequently, many shape judgements can be performed without resolving the relief ambiguity. In the following we shall refer to all shape properties that are preserved by a relief transformation as *relief properties*. A useful and intuitively appealing way of understanding (9) is as an *equivalence class* of three-dimensional shapes. This is well-defined since the relation defines a transformation group.

To summarize the terminology we have introduced, *disparity correction* is the process of computing *scaled relative nearness*, which allows the *relief properties* of the surface to be determined. A graphical illustration of this process is shown in Figure 6.

3.2 Relation to weak calibration models

An important line of research in machine 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).

Some psychophysical results related to these issues have been reported (see e.g. Todd and Norman 1991), but application of the theoretical results to human vision requires some caution, for at least two reasons. First, the equivalence class defined by arbitrary projective (or affine) transformations is almost certainly too general to qualify as a model of “perceptually equivalent shapes” even in a loose sense. Second, the weak calibration assumption does not exploit the fact that the viewing parameters in a fixating binocular vision system are quite constrained. The extrinsic geometry has essentially only three degrees of freedom if Donders’s law is assumed, and it seems reasonable to assume that most of the intrinsic parameters are relatively stable or change only slowly over time.

Hence, it is of interest to study intermediate cases in which the known geometric constraints are exploited. This is precisely what the disparity correction model does, and as expected this makes it possible to compute considerably more structure than in the general projective case.

4 Computational models of vertical disparity processing

In this section we shall consider computational models of how the visual system might use vertical disparities to compensate for the influence of the viewing geometry on the disparity field.

We shall be concerned both with models for *relief reconstruction*, i.e., reconstruction of three-dimensional shape up to a relief transformation, and for *metric reconstruction*, i.e., reconstruction of the full metric structure of the scene. We refer to the process of achieving relief reconstruction from horizontal disparities as *disparity correction* (Section 3). For metric reconstruction it is useful to distinguish between *two stage* processes and *one stage* processes. A two stage process operates according to the theoretical structure outlined above, i.e., by first applying a disparity correction stage to the horizontal disparities and then a *disparity normalization* stage to the output of the first stage. Only the second stage requires explicit knowledge about the viewing geometry. In contrast, a one stage process recovers metric structure directly from horizontal disparities using knowledge of the viewing parameters. Hence, one stage and two stage processes ideally produce the same output, but

in qualitatively different ways. In particular, a two stage model allows vertical disparities to be used in different ways and to different extents in the first and second stages.

We begin by reviewing three different models that have been proposed in the literature. We then explain why neither of these models can satisfactorily account for available psychophysical data, and propose an alternative model called regional disparity correction (RDC).

4.1 The Mayhew and Longuet-Higgins (MLH) model

Mayhew (1982) and Mayhew and Longuet-Higgins (1982) proposed that the viewing parameters needed for one stage metric reconstruction can be estimated from the overall structure of the vertical disparity field. By neglecting the influence of surface structure on the vertical disparity field, which corresponds to setting $1/(d + \delta) \approx 1/d$ (or equivalently, $\lambda \approx \lambda_0$), they noted that the vertical disparity field can be approximated by the simple polynomial function

$$v(x, y) \approx ay + bxy,$$

where a and b are constants. The viewing parameters can then be estimated by

$$d \approx I/b, \quad \gamma \approx a/b.$$

These estimated values can then be substituted back into the horizontal disparity equation (4) to solve for depth. Note that the error caused by the approximation $\lambda \approx \lambda_0$ vanishes for symmetric fixation ($\gamma = 0$). This model, which we shall call the MLH model, assumes zero cyclotorsion, but it was later extended by Porrill *et al.* (1985) to accommodate arbitrary cyclotorsion, e.g. resulting from fixation according to Listing's law.

4.2 The disparity deformation (def) model

The disparity deformation (def) model proposed by Koenderink and van Doorn (1976) performs only disparity correction, and hence computes shape up to a relief transformation.² The def model can be summarized as follows. In a small neighbourhood of the fixation point, the disparity can be expressed as

$$\begin{pmatrix} h(x, y) \\ v(x, y) \end{pmatrix} \approx \begin{pmatrix} h_x & h_y \\ v_x & v_y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad (10)$$

where $(h_x, h_y; v_x, v_y)$ are the gradients of horizontal and vertical disparity respectively, computed at the fixation point ($x = y = 0$). Under the small baseline approximation and assuming zero torsion, (4) and (5) can be used to express this relation as

$$\begin{pmatrix} h(x, y) \\ v(x, y) \end{pmatrix} \approx \frac{I}{d} \begin{pmatrix} P \cos \gamma + \sin \gamma & Q \cos \gamma \\ 0 & \sin \gamma \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad (11)$$

where (P, Q) are the derivatives of the cyclopean depth Z with respect to X and Y . To obtain an expression in terms of scaled relative nearness, we note that

$$\frac{\partial \rho}{\partial x} = I \cos \gamma \frac{P}{d}, \quad \frac{\partial \rho}{\partial y} = I \cos \gamma \frac{Q}{d},$$

²The def model as described by Koenderink and van Doorn actually performs disparity correction with respect to concentric spheres rather than depth planes, but at the fixation point the two are equivalent.

at the fixation point. Hence

$$\begin{pmatrix} h(x, y) \\ v(x, y) \end{pmatrix} \approx \begin{pmatrix} \rho_x + I\lambda_0 \sin \gamma & \rho_y \\ 0 & I\lambda_0 \sin \gamma \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad (12)$$

Any matrix can be expressed as a sum of a symmetric and an anti-symmetric part, and the symmetric part can be further decomposed into one part with vanishing trace and one with non-vanishing trace. The def component of a matrix is defined as the traceless part of the symmetric part; applied to the disparity gradient D' we obtain

$$\text{def } D' = \begin{pmatrix} h_x - v_y & h_y + v_x \\ h_y + v_x & v_y - h_x \end{pmatrix} = \begin{pmatrix} \rho_x & \rho_y \\ \rho_y & -\rho_x \end{pmatrix}. \quad (13)$$

It is now easy to show that the difference of the eigenvalues (ν_1, ν_2) of $\text{def } D'$ satisfies

$$\nu_1 - \nu_2 = \sqrt{\rho_x^2 + \rho_y^2}, \quad (14)$$

and that the eigenvector corresponding to the largest eigenvalue bisects the direction of the gradient (ρ_x, ρ_y) and the projection of the line joining the centers of the left and right eyes onto the cyclopean retina. Hence, the def component of the disparity gradient specifies the gradient $\nabla\rho$ of scaled relative nearness ρ . Moreover, Koenderink and van Doorn (1976) pointed out that this property is invariant with respect to torsional eye rotations. It is also worth noting that since the definition (8) contains an arbitrary constant term, the information content of $\nabla\rho$ is in principle equivalent to that of ρ itself.

4.3 The polar angle disparity (PAD) model

Weinshall (1990) treated the problem of computing a qualitative depth map from the disparity field in the absence of camera calibration information. Rather than decomposing disparity vectors into horizontal and vertical components, Weinshall used a polar decomposition and showed that two different measures derived from the angular component alone contains enough information to compute an approximate depth ordering. Recently, Liu *et al.* (1994) presented numerical simulations showing that the pattern of polar angle disparities can be used to estimate the slope of a planar surface up to scaling by fixation distance, and that this pattern is affected by unilateral vertical magnification in a way which is compatible with Ogle's (1950) induced effect. In this section we shall show that the PAD model can be viewed as a direct and local implementation of disparity correction.

To understand polar angle disparity, consider an observer who is fixating a stationary point in an arbitrarily complex visual scene. Suppose that he moves his head while maintaining fixation in such a way that at the end of the movement the center of each eye is on the same line of sight as it was before, as shown in Figure 7 (left). Since the resulting movement is equivalent to translating each eye independently along its visual axis, the image of any point in the scene is confined to move *radially* from the center of projection in each eye (the focus of expansion). Hence, the disparity component perpendicular to the radial direction (the *angular* component) depends on surface structure and fixation distance, but is invariant to such a gaze shift. (See Figure 7 right.)

For most natural head motions the eyes are not held on precisely the same lines of sight, but it is still true that the angular component of disparity is approximately independent of

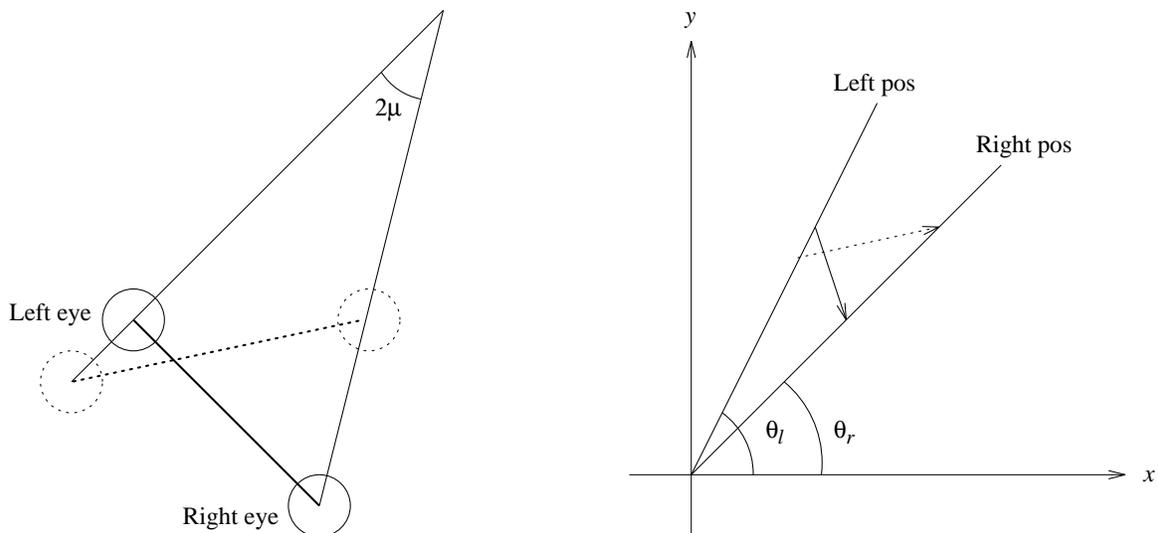


Figure 7: A head motion which changes version γ but preserves vergence μ , resulting in unchanged polar angle disparity. Left: Moving the eyes from the solid circles to the dotted circles along the visual axis of each eye does not affect the vergence angle μ , but does affect the version γ . Right: Because the eye motion is purely translational in the straight-ahead direction, the points projected to either eye are confined to move radially from the fovea. Hence, the difference in polar angle, i.e., the polar angle disparity, remains invariant.

gaze. To show this, we introduce planar polar coordinates (ρ, θ) in the usual way, i.e.,

$$\begin{aligned} x &= \rho \cos \theta, \\ y &= \rho \sin \theta, \end{aligned}$$

for the cyclopean image coordinates, and analogously for the left and right image coordinates. Following Liu *et al.* (1994), we then define the *polar angle disparity* by

$$\Delta\theta = \theta_r - \theta_l. \quad (15)$$

Using the small baseline approximation, it can be shown that

$$\Delta\theta = \frac{hy - vx}{x^2 + y^2}, \quad (16)$$

where (h, v) are the horizontal and vertical disparity defined by (2) and (3).

$\Delta\theta$ is closely related to the quantity

$$\chi = x_r - \frac{y_r}{y_l} x_l,$$

used by Weinshall (1990) to compute a qualitative depth ordering, as can be seen by rewriting

$$\chi = \frac{1}{y_l} (hy_l - vx_l) \approx \frac{1}{y} (hy - vx), \quad (17)$$

where the last equality holds exactly to first order in I/d . Hence, by comparison with (16),

$$\chi = \frac{x^2 + y^2}{y} \Delta\theta \quad (18)$$

Relation to the def model

The def model and the polar angle disparity model are both local in the sense that they do not assume any particular global functional form of the vertical disparity field. However, there are also important differences. First, the polar angle disparity theory produces scaled relative nearness values directly, rather than the gradient of nearness. Second, the polar angle disparity can be computed from a single matched feature, whereas the def theory requires the gradient of disparity to be estimated. As a consequence, the def theory applies specifically to the case of a smooth surface, whereas polar angle disparity in principle can be computed for arbitrarily scattered points. Third, the def theory is invariant to arbitrarily large independent torsional rotations of both eyes, and it does not fail near the horizontal meridian.

It turns out that if we consider the local structure of polar angle disparity instead of individual disparity values, the two theories become almost equivalent. Let

$$a = (x^2 + y^2) \Delta\theta = hy - vx.$$

Using (10) to express h and v in the neighbourhood of the fixation point, we obtain

$$a(x, y) \approx -v_x x^2 + (h_x - v_y)xy + h_y y^2,$$

or equivalently

$$a(x, y) = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} -v_x & \frac{1}{2}(h_x - v_y) \\ \frac{1}{2}(h_x - v_y) & h_y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \quad (20)$$

Denote the matrix in (20) by G . In analogy with def D' , the difference of the eigenvalues of G satisfies (14). Moreover, the eigenvector corresponding to the largest eigenvalue of G bisects the *perpendicular* to the tilt (P, Q) and the projection of the line joining the centers of the left and right eyes onto the cyclopean imaging surface.

It only remains to be shown that these properties are invariant with respect to torsion. First, consider cycloverision; this only results in a corresponding rotation of the cyclopean image coordinates, so that G is transformed into $U^T G U$ where U is an orthogonal rotation matrix. This does not affect the eigenvalues, and the eigenvectors simply follow the rotation of the coordinate system and hence remain invariant relative to the surrounding space. Next, consider cyclotorsion; this means that a constant τ (the cyclotorsion angle) is added to $\Delta\theta$, and hence to both the diagonal elements of G . It is easily verified that this affects neither the eigenvalues nor the eigenvectors of G .

In summary, the local structure of the polar angle disparities determine the gradient of nearness in much the same way as the def component of the disparity gradient.

4.4 The need for an improved model

All three models described above are to some extent inspired by Ogle's (1950) induced effect, i.e., the fact that unilateral vertical magnification causes a fronto-parallel plane to appear slanted. It is therefore not surprising to find that all these models predict exactly the same effect of unilateral vertical magnification, and, moreover, that this prediction agrees well with Ogle's experimental data. Nevertheless, there also exists significant psychophysical evidence *against* each of these models, as will be summarized below.

The def and PAD models both predict that local manipulations of either horizontal or vertical disparity should give rise to the same perceived shape distortions. In practice, however, this symmetry is not observed; local horizontal disparity manipulations give rise to strong and stable effects whereas local vertical disparity manipulations often fail to produce any three-dimensional percept at all. In relation to the def model this asymmetry has been discussed e.g. by Porrill *et al.* (1985). In Appendix B we describe an experiment which leads to a similar conclusion regarding the PAD model. Hence, it appears that *vertical disparity processing is not strictly local*. In the def and PAD models, however, the resolution at which vertical disparities are used is identical to the resolution at which scene structure is computed. Hence, if we try to modify e.g. the def model by postulating that it uses the average deformation of, say, each hemifield, then the resolution of the perceived surface structure would also be on the order of a hemifield, which is clearly absurd. Additional evidence against a pure def theory comes from the observation by Cagenello and Rogers (1990) and Rogers (1992) that binocular image pairs related by a vertical shear produce cyclotorsional eye movements. Such movements would be superfluous if the visual system could extract the deformation component directly.

The MLH model does not distinguish between the influence of vertical disparity on perceived relief structure from that on perceived metric structure. In practice, however, *vertical disparities have only a weak influence on metric tasks*. Sobel and Collett (1991) as well as Cumming *et al.* (1991) have reported that manipulations of the vertical disparity field which simulate a different viewing distance have no effect on the perceived metric shape of objects in the scene, in sharp contrast to the predictions from the MLH model. More recently, Rogers and Bradshaw (1993) have reported an effect obtained with a large field of view, but this effect is still much smaller than that predicted by MLH. In contrast, it appears that visual tasks that only depend on relief structure are strongly influenced by vertical disparity manipulations. Examples are the induced effect (Ogle 1950) and the threads and beads experiment (Helmholtz 1910); a detailed discussion can be found in Section 5.

In light of these observations, we propose below an improved model which combines the desirable properties of previous models while avoiding their main weaknesses. More precisely, this model has the properties that

- like def and PAD, it performs disparity correction based on vertical disparities whereas disparity normalization is left to be performed by a separate mechanism which may or may not use vertical disparities,
- like MLH, it decouples the resolution at which vertical disparities are used from the resolution at which scene structure is computed.

4.5 The regional disparity correction (RDC) model

The MLH model uses the overall structure of the vertical disparity field to compensate for the influence of the viewing geometry on the horizontal disparity field. This idea lends itself even more naturally to a disparity correction process. To see this, it is useful to rewrite the disparity equations (6) and (7) as

$$h(x, y) = \rho(x, y) + xf(x, y), \quad (21)$$

$$v(x, y) = yf(x, y), \quad (22)$$

where

$$f(x, y) = I(\lambda(x, y) \sin \gamma + x \lambda_0 \cos \gamma)$$

The idea, then, is to approximate the vertical disparity field by a simpler polynomial function

$$v(x, y) \approx y \hat{f}(x, y), \quad (23)$$

where

$$\hat{f}(x, y) = a + bx + cy + \dots,$$

which can then be substituted in (21) to obtain the estimate

$$\rho(x, y) \approx h(x, y) - x \hat{f}(x, y) = h(x, y) - x(a + bx + cy + \dots). \quad (24)$$

For example, the MLH approximation $\lambda \sin \gamma \approx \lambda_0 \sin \gamma$ gives

$$f(x, y) \approx I \lambda_0 (\sin \gamma + x \cos \gamma) = a + bx,$$

and hence

$$\rho(x, y) \approx h(x, y) - ax - bx^2. \quad (25)$$

This level of approximation is necessary for metric reconstruction, since explicit estimates of the viewing parameters are then required. For disparity correction, however, there are no such constraints. $f(x, y)$ can then be approximated by a more complex polynomial function to accommodate the variation of the vertical disparity field due to surface structure of arbitrarily high order. For example, for a general planar surface $Z = PX + QY + d$, (6) and (7) can be rewritten as

$$h = I \lambda_0 \left\{ x(P \cos \gamma + \sin \gamma) + yQ \cos \gamma - xyQ \sin \gamma + x^2(\cos \gamma - P \sin \gamma) \right\}, \quad (26)$$

$$v = I \lambda_0 \left\{ y \sin \gamma + xy(\cos \gamma - P \sin \gamma) - y^2Q \sin \gamma \right\}, \quad (27)$$

from which we obtain

$$f(x, y) = I \lambda_0 \{ \sin \gamma + x(\cos \gamma - P \sin \gamma) - yQ \sin \gamma \} = a + bx + cy.$$

Consequently, by fitting a linear three-parameter model to the vertical disparity field, a fixation-invariant estimate of the relief structure is obtained which is unbiased by the overall slope of the surface.

In the preceding analysis we have assumed that the vertical disparity field in the entire field of view is approximated by a single function. This is by no means necessary, however. For example, each hemifield might be approximated by a separate function. For the moment we therefore avoid specifying the precise scale at which information about the viewing geometry is extracted from the vertical disparity field; hence the term *regional* disparity correction. We shall return to the question of scale in Section 6.2.

It may be worth emphasizing that the main justification for the approximation (23) on which the RDC model is based is that the part of the vertical disparity field that depends on surface structure is modulated by the (generally small) factor $\sin \gamma$. Any errors caused by the approximation will thus tend to zero as the viewing geometry approaches symmetric vergence.

Relation to the polar angle disparity model

There is a very close relation between RDC and the polar angle disparity (PAD) model described in Section 4.3. From (19) we have

$$\rho = \chi = h - \frac{x}{y} v. \quad (28)$$

Now suppose that the vertical disparity field is approximated by (23) as in the RDC model. Substituting this relation into (28) we obtain

$$\rho = \chi \approx h - x\hat{f}(x, y),$$

which is precisely the RDC model (24). Hence, PAD can be viewed as the limiting case of RDC in which the scale of the vertical disparity processing goes to zero, i.e., the disparity correction with respect to the viewing geometry is performed independently for each disparity vector. At the other end of the spectrum we have the case in which the disparity correction is based on the properties of the entire vertical disparity field.

5 Vertical disparity and psychophysics

We have shown computationally that metric reconstruction from disparities can be decomposed into disparity correction, which yields three-dimensional shape up to the group of relief transformations, and disparity normalization, which yields full metric structure. A characteristic feature of both of these processes is that neither of them can operate on horizontal disparity alone; unless non-visual cues are used, it is necessary to use the vertical component of disparity as well. The question of if and how the human visual system exploits the information in the vertical disparity field has stimulated several psychophysical studies (Ogle 1950; Stenton, Frisby and Mayhew 1984; Gillam, Chambers and Lawergren 1988; Cumming, Johnston and Parker 1991; Sobel and Collett 1991; Rogers 1992; Rogers and Bradshaw 1993). These studies are based on minor variations of two different ways of manipulating retinal vertical disparities. The first is *unilateral vertical magnification*, i.e., vertical magnification of one eye's view by some factor $1 + \epsilon$. The second is *vertical disparity scaling*, i.e., scaling of the vertical component of each disparity vector by some factor k .

In this section we shall reexamine the results of these studies, in light of the division of psychophysical tasks into on the one hand tasks that can be performed based on relief structure alone, e.g. planar/curved discrimination and depth ordering, and on the other hand tasks that force the subject to estimate metric structure by resolving the relief ambiguity, e.g. depth amplitude judgements. As a result of this way of structuring the problem, a remarkably clear picture will emerge: we shall see that relief tasks exhibit a large and stable dependence on the structure of the vertical disparity field, whereas metric tasks are affected very weakly or not at all.

5.1 Some general observations

Anisotropy of disparity correction. From the basic disparity correction equation (6) it is evident that scaled relative nearness is obtained from horizontal disparity by *adding* a term that depends on horizontal retinal eccentricity x . Hence, if this term is obtained from vertical disparity, we expect a fundamental horizontal/vertical anisotropy in the effect

of any vertical disparity manipulation; for example, there should be no effect at all on the vertical meridian $x = 0$, since scaled relative nearness is equal to horizontal disparity there. In contrast, disparity normalization is inherently isotropic in this sense. It operates on the scaled relative nearness values and computes metric depth by resolving the relief ambiguity (9), which treats the horizontal and vertical dimensions in exactly the same way.

Equivalence of disparity correction models. Since the def, PAD and RDC models of the disparity correction process differ mainly in the scale at which vertical disparity information is used, it is not surprising to find that for global unilateral magnification as well as global vertical disparity scaling, all three models yield exactly the same predictions. Hence, no experiment based on this type of global vertical disparity manipulation can be used to distinguish between these models. As pointed out in Section 4, the evidence against def and PAD comes from experiments based on *local* vertical disparity manipulations.

Symmetric fixation and vertical disparity. In all experiments we consider, the stimulus is viewed with symmetric fixation, i.e., $\gamma = 0$. Under this condition, (6) and (7) simplify to

$$h = \rho + I\lambda_0 x^2, \quad (29)$$

$$v = I\lambda_0 xy. \quad (30)$$

It is worth emphasizing that the scaled relative nearness ρ does not appear in (30); under symmetric fixation and to first order in I/d , the vertical disparity field is in fact *independent* of the depth structure of the scene.

Visualizing relief structure. Visualization of the predicted effect on perceived relief structure by a given disparity manipulation is complicated by the fact that the output of disparity correction is an equivalence class of surfaces rather than a single surface. We shall adopt the convention of always showing the particular member of this class of surfaces that corresponds to the *true* gaze and fixation distance. In other words, this surface is obtained by applying the correct disparity normalization stage to the output of the disparity correction stage, regardless of the disparity manipulation.

5.2 Unilateral vertical magnification: the induced effect

If one eye's view is magnified vertically, a frontoparallel plane appears slanted around a vertical axis towards the eye with the magnifying lens (Ogle 1950). This effect results from disparity correction rather than disparity normalization, and it is predicted by all three computational models we have considered.

The vertical disparity v' resulting from vertical magnification of the right eye's view by a factor $1 + \epsilon$ is

$$v' = (1 + \epsilon)y_r - y_l = v + \epsilon y_r.$$

Ignoring the small difference between left and cyclopean image coordinates and substituting v from (30) for symmetric fixation, the resulting vertical disparity field is

$$v' = \epsilon y + I\lambda_0 xy.$$

For a fronto-parallel plane $\rho = 0$, which when substituted in (29) gives the horizontal disparity field

$$h = I\lambda_0 x^2.$$

The RDC model can in this case represent the vertical disparity field exactly by the polynomial function $v' = yf'(x, y)$ with $f'(x, y) = \epsilon + I\lambda_0 x$. From (24) we obtain the estimated scaled relative nearness $\hat{\rho}$

$$\hat{\rho} = h - xf'(x, y) = -\epsilon x. \quad (31)$$

It is easily verified that the PAD model (19) and the def model (13) yield precisely the same prediction. The effect in terms of the slant at the fixation point of the surface reconstructed using the veridical fixation distance d is given by

$$P' = \frac{\partial Z'}{\partial X'} = \frac{d}{I} \rho_x = -\frac{d}{I} \epsilon.$$

Cross-sections of the surfaces reconstructed from the estimated scaled relative nearness $\hat{\rho} = -\epsilon x$ are shown in Figure 9, for two different fixation distances. Note that as predicted the slant increases in proportion to the viewing distance.

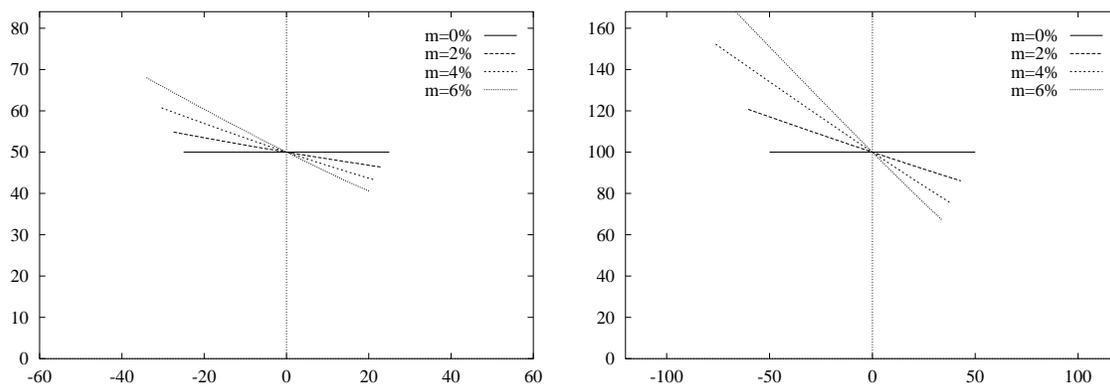


Figure 9: The effect of magnifying the right image $m\%$ vertically and reconstructing the surface by disparity correction followed by veridical disparity normalization, when looking at a fronto-parallel plane. The diagrams show horizontal cross-sections of the reconstructed surfaces; the position of the cyclopean eye is $(0, 0)$. The fixation distance d is 50cm in the left diagram, and 100cm in the right diagram.

The effect of unilateral vertical magnification on the polar angle disparity field has been studied in detail by Liu *et al.* (1994) in a series of numerical simulations. Their results, which were based on an exact representation of the viewing geometry, confirm the predictions derived above which were based on the small baseline approximation. Moreover, Liu *et al.* point out that their results are in good agreement with Ogle's (1950) psychophysical data. This conclusion automatically applies also to the def and RDC models, since, as pointed out above, they yield identical predictions for this type of global vertical disparity manipulation.

5.3 Vertical disparity scaling

This experimental paradigm is based on the idea of replacing the vertical disparity field by the field that would occur at a different viewing distance. Under the small baseline

approximation and symmetric fixation, it is evident from (30) that this is equivalent to scaling the vertical component of every disparity vector by a global factor

$$k = \frac{d}{d'} \quad (32)$$

where d and d' are the true and the simulated fixation distances, respectively.

This type of vertical disparity manipulation is most obviously related to the disparity normalization mechanism, i.e., resolution of the relief ambiguity to obtain metric depth. Resolving the relief transformation (9) near the fixation point is approximately equivalent to scaling horizontal disparity by d^2 and retinal size by d . It is therefore clear that an incorrect estimate of d will distort the perceived depth-shape of an object. If d is estimated from the global vertical disparity field as proposed by Mayhew and Longuet-Higgins (1982), we would hence expect vertical disparity scaling to have a direct effect both on absolute perceived depth and on the ratio between perceived size in the depth and transverse dimensions. This effect appears to be difficult to observe; Sobel and Collett (1991) and Cumming *et al.* (1991) did not find any significant effect of vertical disparity scaling on metric perceptual tasks. More recently, however, Rogers and Bradshaw (1993) have reported a small but significant effect obtained with a large field of view.

However, vertical disparity scaling also affects the disparity correction process. Using (30) to express vertical disparity, we have $v' = kv = yf'(x, y)$ where $f'(x, y) = kI\lambda_0 x$. From the RDC model (24), we obtain the estimate

$$\hat{\rho} = h - xf'(x, y) = \rho + (1 - k)I\lambda_0 x^2. \quad (33)$$

Again, it is easily verified that the PAD model (19) yields the same prediction, and that the def model (13) predicts no effect on the perceived nearness gradient at the fixation point, in keeping with the predictions from the other two models. Hence, vertical disparity scaling should affect disparity correction by *adding* a perceived nearness proportional to the square of the horizontal retinal eccentricity x . This prediction is in fact well supported by psychophysical data, as we shall show in the next two subsections.

Perceived shape of the fronto-parallel plane

Helmholtz (1910) noted that vertical threads in a fronto-parallel plane appear to form a convex surface unless the threads are provided with beads that allow vertical disparities to be detected.

The fact that this effect is anisotropic (the surface appears to be curved around a vertical axis) is an indication that it must originate in the disparity correction process, since, as we have seen, disparity normalization treats the horizontal and vertical dimensions equally. Indeed, assuming that the threads without the beads give rise to zero measured vertical disparities, the equations derived above predict exactly this effect. Figure 10 shows the surfaces reconstructed from (33) when the true surface is a fronto-parallel plane at 50cm, and vertical disparities are scaled by factors $k \in [0, 2]$.

Rogers (1993) has recently reported a large effect (60-70%) of vertical disparity scaling on this task, and proposed that this may be due to the existence of computational “short-cuts” for detecting fronto-parallel surfaces without explicitly estimating the viewing distance. Rogers proposed a mechanism based on ratios of the projected horizontal and vertical sizes of objects in the scene, but alternative and more general models can also

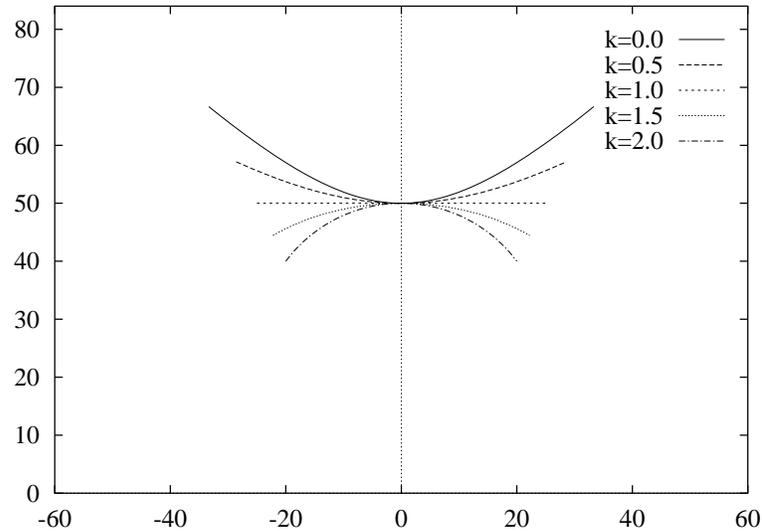


Figure 10: The effect of scaling vertical disparities by a factor k (or equivalently, simulating the viewing distance $d' = d/k$) and reconstructing the surface by disparity correction followed by vertical disparity normalization, when looking at a fronto-parallel plane. The diagram shows a horizontal cross-section of the reconstructed surface; the position of the cyclopean eye is $(0, 0)$. The true fixation distance d is 50cm.

be found; for example, a planar surface perpendicular to the cyclopean line of sight is always characterized by a purely radial pattern of disparities. However, there is no need to postulate the existence of any special-purpose mechanism at all, because a general disparity correction mechanism suffices for any planarity judgement task, without restriction to fronto-parallel planes.

Shape judgements and the horizontal/vertical anisotropy

Cumming *et al.* (1991) investigated the effect of vertical disparity scaling on a metric shape judgement task, illustrated in Figure 11. To perform this task, the subjects must in principle first estimate the vertical extent of the cylinder by scaling the size of its retinal projection by d and then estimate the extent in depth of the cylinder by scaling the disparity by d^2 . Clearly, this could not be done using scaled relative nearness alone, since the relief transformation (9) does not preserve the ratio between the depth and transverse dimensions (see Figure 5). The result of the experiment was clear-cut; there was no significant effect of vertical disparity scaling on the perceived shape of the cross-section. This appears to be a strong indication that the subjects did not use the vertical disparity pattern to estimate d , either directly in a one stage metric reconstruction process or for disparity normalization in a two stage process.

In contrast, this experiment has no bearing at all on models of the disparity correction process. Since vertical disparity scaling was shown above to *add* a vertical cylindrical function to the perceived scaled relative nearness (see Figure 10), the effect is to bend the cylinder around a vertical axis, creating a saddle-shape surface when vertical disparities are increased and a rounded convex hill when they are decreased. In neither case will the shape

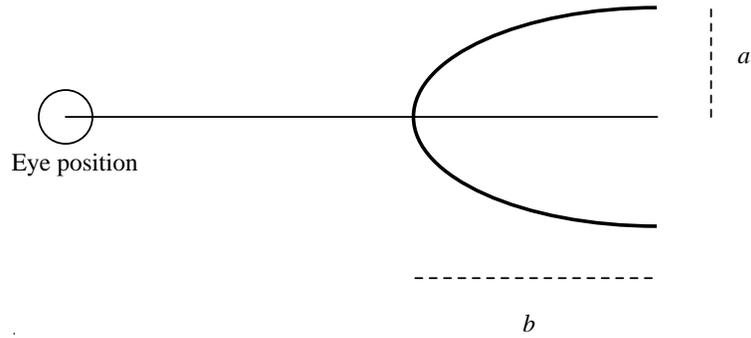


Figure 11: The shape judgement task of Cumming et al. The subjects were shown frontal views of horizontal elliptical cylinders with cross-sections varying from flattened to elongated ellipses, and were asked to determine when the cross-section was circular, i.e., when a appears to be equal to b .

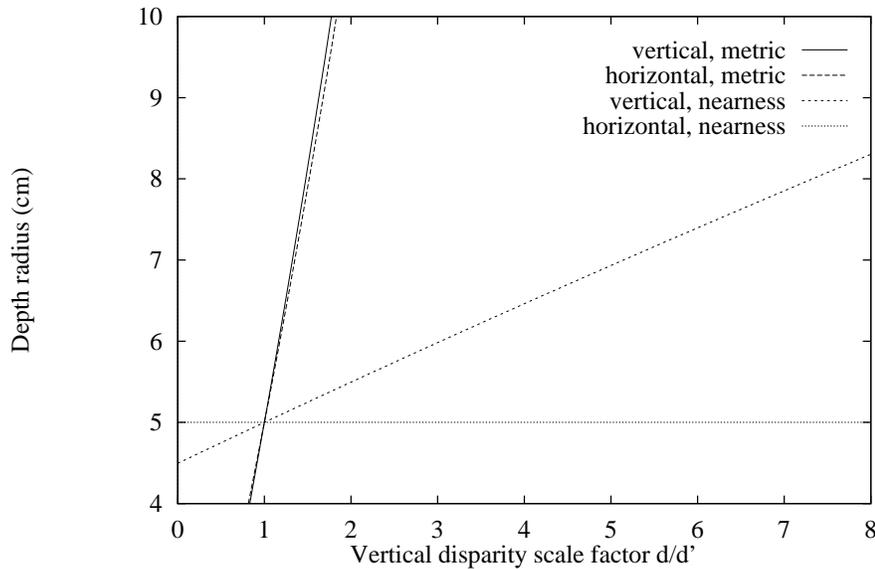


Figure 12: The predicted effect of vertical disparity scaling on the perceived shape of horizontal and vertical elliptical cylinders under two different assumptions about the use of vertical disparities; first, that they are only used for relief reconstruction, i.e., for disparity correction but not for disparity normalization, and second, that they are used for full metric reconstruction, i.e., for both disparity correction and normalization. The horizontal axis shows the vertical disparity scale factor, i.e., the ratio of true to simulated fixation distance. The vertical axis shows the depth radius (in cm) of the elliptical cylinder that appears to have a circular cross-section. The transverse radius of all cylinders is 5cm.

of a vertical cross-section of the cylinder be affected. Using random dot stereograms, we have informally verified that this bending effect does indeed occur.³

However, the situation is quite different for *vertical* cylinders; the effect of the vertical disparity scaling then adds to the curvature of the cylinder, significantly affecting the perceived depth difference between near and far points on a horizontal cross-section of the cylinder. To derive an exact prediction for this effect, we make the simplifying assumption that the shape judgement is done by comparing the perceived depth (relative to the fixation point) at the base of the central cross-section to the perceived angular size of the same cross-section. Furthermore, we assume that scaled relative nearness is computed according to (33), but that the disparity normalization mechanism is unaffected by the vertical disparity manipulations, as indicated by the results of Cumming *et al.* (1991). The resulting predictions are shown in Figure 12, for both horizontal and vertical cylinders with 5cm radius viewed at a distance of $d = 50\text{cm}$. The predictions for the case in which vertical disparities are also used for resolving the relief ambiguity to obtain metric shape are shown for comparison.

We are currently conducting a series of experiments (to be reported in a subsequent publication) aimed at testing this horizontal/vertical anisotropy predicted by the disparity correction theory. Initial results from a pilot study (Gårding *et al.* 1993) do indeed appear to confirm the predictions quite well; there was no significant effect on horizontal cylinders, whereas the effect on vertical cylinders was about 65% of that predicted by full disparity correction.

5.4 Vertical disparity and psychophysics: Summary

A summary of the psychophysical data described above is shown in Table 1. It reveals a rather striking difference between relief tasks and metric tasks regarding the influence of vertical disparity.

Task	Effect	Reference
<i>Relief tasks</i>		
Induced effect	100%	Ogle 1950
Fronto-parallel plane	70%	Helmholtz 1910; Rogers <i>et al.</i> 1993
Vertical cylinders	65%	Gårding <i>et al.</i> 1993
<i>Metric tasks</i>		
Step amplitude	0%	Sobel and Collett 1991
Horizontal cylinders	0%	Cumming <i>et al.</i> 1991
Grating amplitude	20%	Rogers <i>et al.</i> 1993

Table 1: Summary of available psychophysical data regarding the effect of vertical disparity manipulations. The strength of the effect is consistently higher for tasks that can be performed from relief structure than for tasks that require estimation of metric structure. The precise percentages shown here should not be taken too literally; their values depend on how the task is defined and on the experimental conditions.

³As pointed out by one of the reviewers, this kind of demonstration requires enormous care, since an incorrect geometry in the stereo viewing apparatus may give rise to a similar effect.

The disparity normalization results reported by Rogers *et al.* (1993) emerge from this breakdown as a very important finding; without these results one would have no reason to believe that disparity normalization could be directly driven by (vertical) disparity.

6 Summary and discussion

6.1 Summary

In the first part of the paper we established a general framework for the analysis of stereopsis under variable viewing geometry. The key feature of this framework is the distinction between on the one hand *disparity correction*, which is used to compute three-dimensional structure up to a relief transformation, and on the other hand *disparity normalization*, which is used to resolve the relief ambiguity to obtain metric structure. The relief transformation was analyzed in further detail, and it was shown that it preserves many important properties of visual shape, notably the depth order as well as all projective properties such as coplanarity and collinearity. As a consequence, the visual system may not very often have to perform the disparity normalization step at all.

In the second part of the paper, we reviewed and analyzed three previously proposed computational models of how the visual system might use properties of the disparity field itself to compensate for the influence of the viewing geometry on horizontal disparities. We observed that there exists significant psychophysical evidence against all of these models, notably (i) that vertical disparity processing is not strictly local, and (ii) that vertical disparities appear to be used to a much larger extent for disparity correction than for disparity normalization. This led us to propose the new *regional disparity correction* (RDC) model, which tries to combine the strengths of all three of the previous models while eliminating their main drawbacks. In light of the RDC model we reexamined psychophysical data concerning the influence of vertical disparity on stereopsis, and derived predictions for some new experiments.

6.2 Discussion

Architecture of stereopsis

Since it appears that vertical disparity is used to a much higher degree for disparity correction than for disparity normalization, it is interesting to ask how other cues than vertical disparity influence these mechanisms. Based on the fact that relief structure is sufficient for many (if not most) perceptual tasks, one may hypothesize that disparity normalization is only performed in the relatively rare cases when recovery of metric structure is essential to the visual task at hand. It may then be performed at a high and more cognitive level of visual processing, in which case we would expect it to be strongly influenced by high-level pictorial cues such as known object size, texture gradients, shading etc. In fact, Glennerster *et al.* (1993) have reported nearly perfect metric depth constancy under full-cue conditions, i.e., when all cues from binocular disparity, vergence angle, accommodation and texture gradients etc. provide consistent information.

The obvious way in which pictorial cues can be used for disparity normalization is by providing an estimate of fixation distance, which is then used for resolving the relief transformation. Theoretically, however, more indirect mechanisms could also be envisaged; for example, an estimate of surface orientation provided by a shape-from-texture module

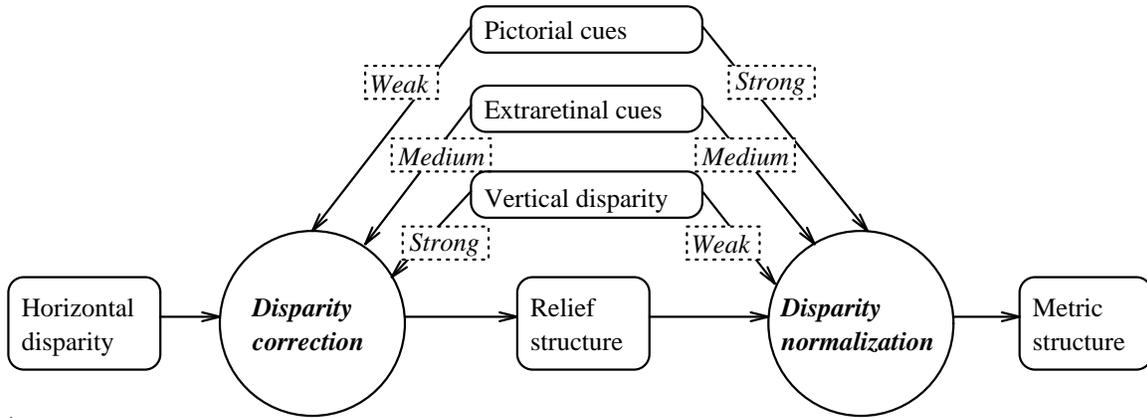


Figure 13: *Speculative block diagram of stereopsis. Some of the indicated interaction strengths remain to be tested experimentally.*

could be fed into (26) and (27) to solve for fixation distance and gaze (Porrill *et al.* 1991); see also (Johnston, Cumming and Parker 1993). However, psychophysical experiments (Frisby *et al.* 1994) based on the judged amplitude of steps set into planar surfaces have so far failed to confirm the existence of such a mechanism in the human visual system.

The extraretinal cues, i.e., feedback from the oculo-motor and/or accommodation systems, appear to take a position intermediate to vertical disparity and pictorial cues. For the perceived fronto-parallel plane, Helmholtz (1910) found (using vertical threads with or without beads as described in Section 5.3) that small changes in vergence angle created with prisms had little or no effect when the stimulus allowed both horizontal and vertical disparities to be estimated, but when the vertical disparity information was removed an effect of the prisms became manifest. Similarly, Rogers (1993) has reported an observed effectiveness of the vergence cue in the range from about 90% for a small field of view to about 20% for a very large field of view, whereas the effectiveness of the vertical disparity cue had roughly the opposite dependence on the size of the field of view. This makes sense since the magnitude of vertical disparities increases with retinal eccentricity according to (30); it indicates that stereopsis operates in a highly adaptive way, attempting to make the best possible use of the available information.

The perceived fronto-parallel plane is a relief task, but several results concerning the effect of the vergence cue on metric tasks have also been reported. For slant judgements, Frisby and Buckley (1992) found an effect of about 10–15%; for the perceived shape of horizontal elliptical cylinders (see Section 5.3), Cumming *et al.* (1991) found an effect of similar magnitude;⁴ and for the perceived amplitude of sinusoidal gratings, Rogers (1993) found an effect in the range 20–35%. Hence, the largest effect is found for a relief task, although the difference between relief tasks and metric tasks seems less pronounced for the vergence cue than for the vertical disparity cue.

These considerations are summarized as a block diagram of stereopsis shown in Figure 13. This diagram is speculative by nature; in particular, the effectiveness of pictorial cues remain to be investigated experimentally, and must also be expected to vary with the experimental conditions.

⁴Cumming *et al.* (1991) assumed that the vergence angle also influences the scaling of retinal size, and estimated the total effect to be 25%.

A final point about the perceptual consequences of omitting the disparity normalization step is in order. As we have shown, the end result of disparity correction is an equivalence class of shapes. However, even if the visual system stops after this stage there does not necessarily exist a corresponding subjective perceptual ambiguity, because the visual system may still represent the stimuli unambiguously by choosing one particular representative of the equivalence class. In our terminology this corresponds to performing disparity normalization with some “default” set of viewing parameters, which may or may not reflect the actual viewing geometry. This representative may be quite stable in the sense that the same disparity pattern always gives rise to more or less the same perceptual interpretation.

The scale of vertical disparity processing

As discussed in Section 4.5 where the RDC model was derived, the scale at which the human visual system processes vertical disparities is not yet exactly determined. However, there exist a number of reported results which can provide upper and lower bounds. If a global representation of the viewing geometry exists, it must be different from that used for representing felt eye position, since vertical disparity manipulations that cause perceptual effects consistent with an incorrectly estimated asymmetric gaze angle do not make observers feel that they are “not looking straight ahead” (Frisby 1984). Moreover, Rogers and Koenderink (1986) observed that unilateral vertical magnification with opposite signs in the left and right halves of the visual field causes subjects to perceive simultaneous and opposite induced slants. This indicates that the mechanism is not strictly global, but half the visual field is still compatible with a highly non-local mechanism. A negative result for the other end of the scale spectrum was reported by Porrill *et al.* (1985), who observed that a stereogram for which a mechanism based on purely local computation of deformation would produce a saddle-shaped percept was not found to produce any stereoscopic depth impression at all. A similar result related to the polar angle disparity (PAD) model is described in Appendix B. See also (Westheimer and Pettet 1992).

Given the computational advantages of local approaches such as def and PAD, one may then ask why the visual system does not appear to have implemented them. One plausible explanation is that the disparity interpretation process is linked to the matching process in which retinal disparities are estimated. As pointed out in the introduction, any efficient implementation of the matching process must be based on the epipolar constraint, which reduces the dimensionality of the search space from two to one. While this constraint can be economically represented by the global parameters of the viewing geometry, other more distributed representations may be more useful in practice. For example, a direct representation of the epipolar line corresponding to each position in one retina for a given viewing geometry would change slowly with retinal position as shown in Figure 4. A consistency-maintaining mechanism in such a representation could involve local interactions that inhibit mutually incompatible epipolar lines in neighbouring regions, while allowing gradual changes such as would occur in most natural fixation positions. An implementation of a neural network architecture in this spirit has been described by Mayhew *et al.* (1992).

A The disparity equations

A straightforward if somewhat tedious way of deriving the disparity equations (4) and (5) under the small baseline approximation is to start from the exact viewing geometry defined in Figure 1 and then differentiate the perspective projection equations with respect to I/d . A shortcut is provided by the observation that the small baseline approximation makes the binocular geometry exactly equivalent to the geometry of image flow. Hence, each disparity vector is proportional to the velocity by which an image point moves when the left eye translates and rotates to occupy the position of the right eye.

Let $\mathbf{r} = (X, Y, Z)^T$. Under rigid body motion the velocity of \mathbf{r} can be expressed as

$$\dot{\mathbf{r}} = \mathbf{t} + \boldsymbol{\omega} \times \mathbf{r},$$

where \mathbf{t} is the instantaneous translation and $\boldsymbol{\omega}$ is the instantaneous rotation. Under perspective projection, the (suitably normalized) cyclopean image coordinates are $x = X/Z$ and $y = Y/Z$, where $Z = d + \delta$ as shown in Figure 1. Differentiating with respect to time and rearranging terms, we obtain the standard image flow equations

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} (t_x - x t_z)/(d + \delta) + \omega_y(1 + x^2) - \omega_z y - \omega_x x y \\ (t_y - y t_z)/(d + \delta) - \omega_x(1 + y^2) + \omega_z x + \omega_y x y \end{pmatrix}. \quad (34)$$

The rate of change of eye-centered coordinates of any fixed point in the world as the eye moves from the left to the right position is equal to the change in fixed coordinates caused by the opposite movement of that point. Hence, we need the translation and rotation that take the right eye to the left eye, and referring to Figure 1 we have

$$\mathbf{t} = -I \begin{pmatrix} \cos \gamma \\ 0 \\ \sin \gamma \end{pmatrix}, \quad (35)$$

and

$$\boldsymbol{\omega} = \begin{pmatrix} 0 \\ 2\mu \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ (I/d) \cos \gamma \\ 0 \end{pmatrix}, \quad (36)$$

where the last equality holds to first order in I/d , as can be seen by differentiating (1). Substituting (35) and (36) into (34), we then obtain the disparity equations (4) and (5).

B The PAD model: A psychophysical experiment

We have performed a simple psychophysical experiment aimed at testing the capability of the human visual system to process vertical disparity in a strictly local way according to the polar angle disparity (PAD) model described in Section 4.3. The stimulus, schematically illustrated in Figure 14, was a random-dot stereogram in which the background is consistent with a fronto-parallel plane. In the two circular regions indicated in the figure, the polar angle disparities were modified from the background plane to define two Gaussian hills in depth according to (19). Since polar angle disparities result from combinations of horizontal and vertical disparities according to (16), this can be achieved either by modifying horizontal disparities and leaving the vertical disparities as they are, or vice versa,

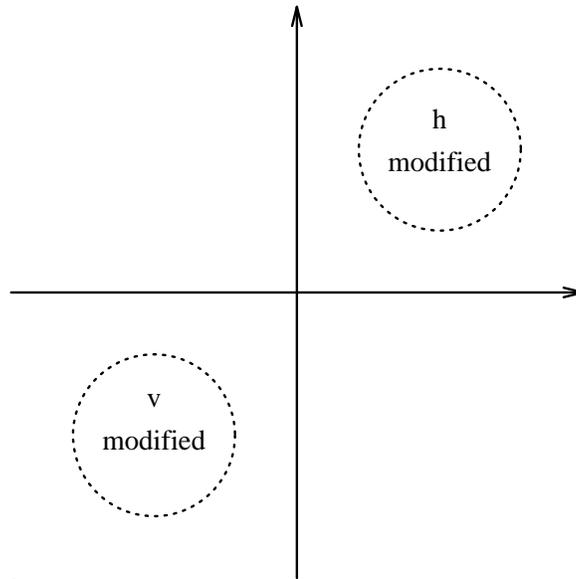


Figure 14: Local manipulation of horizontal and vertical disparity used for testing the polar angle disparity model.

or by any combination of these two methods. In our experiment, one of the Gaussian hills (the h region) was created by modifying only the horizontal disparity, and the other (the v region) by modifying only the vertical disparity. For a computational mechanism that only uses the polar angle component, the three-dimensional percepts should be identical for both regions. However, this is not what the subjects perceived; every subject saw a hill in depth in the h region, whereas the v region was sometimes perceived as rivalrous but did not produce any clear depth percept. Although this experiment was performed in an informal manner, the result was so clear-cut that it seems safe to rule out purely local processing of polar angle disparity as a model of stereopsis in the human visual system.

References

- J.A. Breysig, *Versuch einer Erläuterung der Reliefperspektive*. Magdeburg, 1798.
- R. Cagenello and B.J. Rogers, “Orientation disparity, cyclotorsion and the perception of surface slant”, *Investigative Ophthalmology and Visual Science*, vol. 31, p. 97, 1990.
- R.H.S. Carpenter, *Movements of the Eyes*. Pion Limited, London, second ed., 1988.
- T.S. Collett, U. Schwartz, and E.C. Sobel, “The interaction of oculomotor cues and stimulus size in stereoscopic depth constancy”, *Perception*, vol. 20, pp. 733–754, 1991.
- B.G. Cumming, E.B. Johnston, and A.J. Parker, “Vertical disparities and perception of three-dimensional shape”, *Nature*, vol. 349, pp. 411–413, Jan. 1991.
- 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.M. Foley, “Binocular distance perception”, *Psychological Review*, vol. 87, pp. 411–435, 1980.
- J.M. Foley, “Binocular distance perception: egocentric visual task”, *J. of Experimental Psychology: Human Perception and Performance*, vol. 11, pp. 133–149, 1985.

- J.P. Frisby, "An old illusion and a new theory of stereoscopic depth perception", *Nature*, vol. 307, pp. 592–593, Feb. 1984.
- J.P. Frisby and D. Buckley, "Experiments on stereo and texture cue combination in human vision using quasi-natural viewing", in *Artificial and Biological Vision Systems* (G. Orban and H.H. Nagel, eds.), ESPRIT Basic Research Series, Springer-Verlag, 1992.
- J.P. Frisby, D. Buckley, K.A. Wishart, J. Porrill, J. Gårding, and J.E.W. Mayhew, "Interaction of stereo and texture cues in the perception of three-dimensional steps", *Vision Research*, 1994. (To appear).
- J. Gårding, J. Porrill, D. Buckley, J.P. Frisby, S.D. Hippisley-Cox, J.E.W. Mayhew, and J-O Eklundh, "Binocular shape perception from non-radial disparities", presented at the *NATO Advanced Workshop on Binocular Stereopsis and Optical Flow*, Toronto, Canada, June 1993.
- B. Gillam, D. Chambers, and B. Lawergren, "The role of vertical disparity in the scaling of stereoscopic depth perception: An empirical and theoretical study", *Perception & Psychophysics*, vol. 44, no. 5, pp. 473–483, 1988.
- A. Glennerster, B.J. Rogers, and M.F. Bradshaw, "The constancy of depth and surface shape for stereoscopic surfaces under more naturalistic viewing conditions", *Perception*, vol. 22 (suppl.), p. 118, 1993.
- H.L.F. von Helmholtz, *Treatise on Physiological Optics*, vol. 3. (trans. J.P.C Southall, Dover, New York 1962), 1910.
- B.K.P. Horn, "Relative orientation", *Int. J. of Computer Vision*, vol. 4, pp. 59–78, 1990.
- E.B. Johnston, "Systematic distortions of shape from stereopsis", *Vision Research*, vol. 31, pp. 1351–1360, 1991.
- E.B. Johnston, B.G. Cumming, and A.J. Parker, "Integration of depth modules: Stereopsis and texture", *Vision Research*, vol. 33, pp. 813–826, 1993.
- J.J. Koenderink, "Optic flow", *Vision Research*, vol. 26, no. 1, pp. 161–180, 1986.
- 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.
- L. Liu, S.B. Stevenson, and C.M. Schor, "A polar coordinate system for describing binocular disparity", *Vision Research*, 1994. (To appear).
- J.E.W. Mayhew, "The interpretation of stereo-disparity information: the computation of surface orientation and depth", *Perception*, vol. 11, pp. 387–403, 1982.
- J.E.W. Mayhew and H.C. Longuet-Higgins, "A computational model of binocular depth perception", *Nature*, vol. 297, pp. 376–378, 1982.
- J.E.W. Mayhew, Y. Zheng, and S. Cornell, "The adaptive control of a four-degrees-of-freedom stereo camera head", *Phil. Trans. Royal Society London B*, vol. 337, pp. 315–326, 1992.
- K.N. Ogle, *Researches in Binocular Vision*. Saunders, Philadelphia, 1950.
- A.P. Petrov, "A geometrical explanation of the induced size effect", *Vision Research*, vol. 20, pp. 409–413, 1980.
- J. Porrill, J. Gårding, J.O. Eklundh, J.P. Frisby, D. Buckley, S. Pollard, J. Mayhew, and E. Spivey, "Using shape-from-texture to calibrate stereo", *Perception*, vol. 20, no. 1, p. 90, 1991.
- J. Porrill, J.E.W. Mayhew, and J.P. Frisby, "Cyclotorsion, conformal invariance, and induced effects in stereoscopic vision", Tech. Rep. AIVRU-007, AI Vision Research Unit, Sheffield University, 1985.

- B.J. Rogers, "The perception and representation of depth and slant in stereoscopic surfaces", in *Artificial and Biological Vision Systems* (G. Orban and H.H. Nagel, eds.), ESPRIT Basic Research Series, Springer-Verlag, 1992.
- B.J. Rogers, "The concept of binocular disparity", presented at *NATO Workshop on Binocular Stereopsis and Optical Flow*, Toronto, Canada, June 1993.
- B.J. Rogers and M.F. Bradshaw, "Vertical disparities, differential perspective and binocular stereopsis", *Nature*, vol. 361, pp. 253–255, Jan. 1993.
- B.J. Rogers and J.J. Koenderink, "Monocular aniseikonia: a motion parallax analogue of the disparity-induced effect", *Nature*, vol. 322, pp. 62–63, 1986.
- E.C. Sobel and T.S. Collett, "Does vertical disparity scale the perception of stereoscopic depth?", *Proc. Royal Society London B*, vol. 244, pp. 87–90, 1991.
- S.P. Stenton, J.P. Frisby, and J.E.W. Mayhew, "Vertical disparity pooling and the induced effect", *Nature*, vol. 309, pp. 622–623, June 1984.
- J.T. Todd and J.F. Norman, "The visual perception of smoothly curved surfaces from minimal apparent motion sequences", *Perception & Psychophysics*, vol. 50, pp. 509–523, 1991.
- D. Weinshall, "Qualitative depth from stereo, with applications", *Computer Vision, Graphics, and Image Processing*, vol. 49, pp. 222–241, 1990.
- G. Westheimer and M.W. Pettet, "Detection and processing of vertical disparity by the human observer", *Proc. Royal Society London B*, vol. 250, pp. 243–247, 1992.