

A psychophysical and computational analysis of intensity-based stereo

Hanspeter A. Mallot¹, Petra A. Arndt², Heinrich H. Bülthoff¹

¹ Max-Planck-Institut für biologische Kybernetik, Spemannstrasse 38, D-72076 Tübingen, Germany

² Institut für Kognitionsforschung, Carl von Ossietzky Universität, D-26111 Oldenburg, Germany

Received: 9 June 1995 / Accepted in revised form: 3 June 1996

Abstract. We describe two psychophysical experiments testing predictions of the square difference mechanism we have previously proposed for intensity-based stereo. Experiment 1 assesses the relative contributions of disparity and contrast to intensity-based stereo by measuring detection thresholds. The product of disparity and contrast at threshold is shown to be constant. In experiment 2, we measure quantitatively the global depth position perceived in stereograms of curved, smoothly shaded surfaces. The results show that disparity averaging over the surface involves a contrast-dependent weighting function. The results from both experiments are consistent with predictions derived from the square difference mechanism. The relation of this mechanism to feature correspondence stereopsis and shape-from-shading is discussed and a general framework for assessing the modularity of stereopsis is presented.

1 Introduction

1.1 Disparity tuning and correspondence

Theories of human stereoscopic depth perception can be broadly classified into two groups: a neurobiological approach based on disparity-tuned units and a computational approach focusing on the stereo correspondence problem. Starting from experimental findings in humans (Richards 1971), cats (Bishop et al. 1971) and monkeys (Poggio and Fischer 1977), disparity-selective units have been studied in great detail. For recent reviews see Poggio (1995), DeAngelis et al. (1995), and Howard and Rogers (1995). Disparity selectivity is thought to be constructed in two steps by simple and complex cells. Binocular simple cells have two linear receptive fields, one in the left and one in the right eye, which differ in spatial position or Fourier phase. These differences between the left and right receptive fields provide the initial data from which disparity estimates are obtained. In the second step, binocular complex cells sum the squared responses of a number of simple cells (Ohzawa et al. 1990;

Qian 1994). By combining the complex cell excitations in a population code (e.g., Lehky and Sejnowski 1990), simple stereo tasks such as global stereopsis (Julesz 1971) or vergence control (Mallot et al. 1996b) can be modelled.

The computational view on stereo focuses on the stereo correspondence problem, i.e., the identification of pairs of feature points (or regions) in the left and right image that depict the same object in the outside world (Julesz 1971; Marr and Poggio 1979). Demonstrations of the correspondence problem include the well-known double-nail illusion (Krol and van de Grind 1980) as well as the wallpaper illusion (for references see Mallot et al. 1996a; Howard and Rogers 1995). The correspondence problem is particularly difficult if stimuli with many similar features are used. In computer vision applications, it is the price to be paid for discarding most of the gray-level information in the initial image by restricting it to a primal sketch, i.e., to a list of localized feature points. Disparity-selective units *per se* do not solve the stereo correspondence problem. However, if lateral interactions in an array of disparity-tuned units are considered, false matches can be suppressed and unambiguous solutions of the correspondence problem can be obtained (Julesz 1971; Marr and Poggio 1976; for a review of these and other ‘cooperative’ stereo theories, see Blake and Wilson 1991).

In summary, three steps in a neural theory of stereopsis can be distinguished:

1. Unocular preprocessing is described by the left and right receptive fields of binocular simple cells. The differences between these receptive fields define the raw disparity data (feature locations, position or Fourier phase shift of gray-level patches) that subsequent steps have to rely on.
2. Binocular interaction involves nonlinearities such as the computation of ‘disparity energy’ by taking sums of squares of the simple cell outputs. The result of this stage is coded by the excitation of the disparity-tuned units.
3. Cooperation or lateral interaction between disparity-tuned units can solve the stereo correspondence problem and thus provide high-resolution depth maps of curved or corrugated surfaces.

In this paper, we present psychophysical data relevant to the second stage, specifically the nonlinearity of the binocular interaction. We will show that quadratic nonlinearities quantitatively explain the results from two experiments concerning threshold and depth averaging in intensity-based stereo.

1.2 Intensity-based stereo

In recent experiments with smoothly shaded gray wedges, we have shown that stereo disparities in continuous gray-level images can be detected even though salient image features were missing and, consequently, point disparities could not be defined (Arndt et al. 1995). In a systematic test of a number of potential candidates for feature matching, we showed that neither luminance edges nor luminance extrema nor the overall centroid of a gray wedge are necessary to perceive stereoscopic depth. Rather, the differences in luminance profiles of the left and right eye's images suffice. These results confirm and extend earlier findings on the role of contrast disparities in stereopsis (Westheimer and McKee 1980; Mayhew and Frisby 1981; Bülthoff and Mallot 1988; Christou and Parker 1993).

As an underlying mechanism, Arndt et al. (1995) suggested the minimization of the mean square difference of the image profiles by varying the horizontal offset between the images. This mechanism, which is equivalent to maximizing area correlation, is well in line with psychophysical experiments and with computational models of disparity-tuned units (see above). Cormack et al. (1991), for example, found that stereo acuity decreases with interocular correlation in random dot stereograms superimposed with uncorrelated noise. With ambiguous random dot stereograms made of multiple 'double-nail' patterns, Weinshall (1991) showed that 'ghost' planes can be seen at depth positions that correspond to peaks in the correlation function. The ghosts do not appear in stimuli with only one or a few double-nail patterns, indicating that this correlation-based percept is related to global stereopsis (Julesz 1971). Note that a stereo mechanism using interocular correlation or squared image difference would not respond to mere contrast differences without positional shifts (disparities). As was shown by Blake and Cormack (1979), no stereoscopic depth perception results from pure contrast disparities in sine wave gratings.

1.3 Predictions of the square-difference mechanism

In this paper, two predictions of the square difference mechanism for intensity-based stereo are tested quantitatively. Let us first state the proposed theory in mathematical terms. We consider only intensity profiles whose spatial extent is small with respect to the correlation window; minimization of the square difference thus leads to just one global disparity value for the entire stimulus. For a stereogram composed of two half-images $I_l(x, y)$, $I_r(x, y)$, consider the space-independent difference function

$$\Phi(D) = \int \int |I_l(x, y) - I_r(x + D, y)|^2 dx dy \quad (1)$$

The disparity estimate D is obtained by minimizing Φ :

$$\Phi(D) = \min \Phi(D) \quad (2)$$

From these equations, we derive predictions concerning the relation of disparity- and contrast-threshold for the perception of depth order as well as the amount of depth perceived in intensity-based stereo:

Prediction 1 (Threshold). Shifts of the stereogram off the zero-disparity plane should be perceived, if the squared image difference (1) exceeds a threshold: $\Phi(0) > \Phi_o$. The size of $\Phi(0)$ depends on both, image contrast and disparity. Consider a stereogram with constant disparity δ and intensity amplitude c : $I_l(x) = cI(x - \delta/2)$, $I_r(x) = cI(x + \delta/2)$. From (1), we have:

$$\Phi(0) = \int \int |cI(x - \delta/2, y) - I(x + \delta/2, y)|^2 dx dy \quad (3)$$

We expand the difference into a Taylor series and note that all terms with even-numbered derivatives cancel out. As a second-order approximation, we obtain

$$\Phi(0) \int \int c^2 \delta^2 \left(\frac{\partial I(x, y)}{\partial x} \right)^2 dx dy \quad (4)$$

$$c^2 \delta^2. \quad (5)$$

At threshold, we thus obtain $\Phi_o = c^2 \delta^2$ and further $c \propto \delta^{-1}$. In other words, contrast and disparity should be reciprocally related to each other at threshold; e.g., if contrast is halved, disparity must be doubled to restore visibility. In experiment 1, this prediction is tested using the parabolic gray wedge introduced by Arndt et al. (1995).

Prediction 2 (Depth shift). In stimuli with different disparities (i.e., images of slanted or curved surfaces), intensity-based stereo tends to produce one global depth estimate rather than a well-resolved depth map of the surface. If $\delta(x, y)$ denotes the true disparity distribution, the global disparity estimate minimizing the square difference can be shown to be:

$$D = \frac{\int \int \delta(x, y) I^2(x, y) dx dy}{\int \int I^2(x, y) dx dy} \quad (6)$$

where I denotes the partial derivative of the image function with respect to x (see Appendix A and Eq. 15 of Arndt et al. 1995). D is an average of the true disparities weighted with I^2 , i.e., a measure of local image contrast in the x -direction. In experiment 2, we test this prediction with ortho- and pseudoscopic stereograms of smoothly shaded ellipsoids of various elongations.

In addition, both experiments reveal a number of differences between intensity-based and local, feature-based stereopsis. In Sect. 5 the modularity of stereopsis will be discussed against the background of these (and other) differences.

2 General methods

2.1 Subjects

Six volunteers, aged 23–28 years, participated in the reported experiments. All but two subjects (U.B. and M.J.) were experienced observers of experiments on stereo vision. All subjects passed a random dot stereogram test for binocular vision. If necessary, vision was corrected with spectacles during experimental sessions.

2.2 Stimulus presentation.

Stereograms were presented on a color display monitor (Mitsubishi Color Display Model no. HL6905 STGR) in interlaced mode, each half-image with a frequency of 60 Hz. The presentation of the half-images to the left and right eye was controlled by liquid crystal shutter glasses (Stereographics; cf. Hodges 1992). The average transmittance of the glasses was 13% and peak transmittance in the open phase was 30%.

The luminances produced by the monitor for 52 of its 256 color map slots were measured with a UDT optometer. The resulting calibration curve reflects the built-in γ -correction of the monitor (cf. Foley et al. 1990). For any desired luminance, the required color map slot of the monitor (0–255) was determined by inversely evaluating the calibration curve.

One- and two-dimensional smooth intensity profiles subtending 6×6 deg of visual angle were used as stimuli. Stimuli are described in detail below. All experiments were performed under dark room conditions. The viewing distance was 115 cm and the head of the subject was fixed with a forehead and chin rest. Two stereograms made of the same pair of half-images in ortho- or pseudoscopic order were displayed simultaneously on the monitor.

3 Experiment 1. Contrast and disparity at threshold

3.1 Stimulus

Stimuli lacking Laplacian zero- and level-crossings as well as other features have been discussed at length in Arndt et al. (1995). The simplest one is a parabolic intensity profile, the second derivative of which is constant. We used parabolic luminance profiles with intensity maxima in the center of the stimulus for threshold measurements. The corresponding equations for the two half-images read:

$$I_a(x) := M + A \frac{(1 + \delta_o) + \delta_o x - x^2}{(1 + \delta_o/2)^2}; \quad \frac{\delta_o}{2} \in [0, 1]$$

$$I(x) := I_a(-x) \quad (7)$$

Here, $\delta_o/2$ is the peak position of the parabola, i.e., δ_o is the presented disparity. Stimulus contrast was defined as two-point contrast between the brightest and the darkest point in the intensity profile, i.e., $c = A/(2M + A)$, (Fig. 1a,b). Contrast changes were carried out in such a way that the mean intensity of the profile was kept constant.

The half-images I_a and I were combined in two different stereograms S_a (half-image I_a presented to the left eye and half-image I to the right eye) and S_a (vice versa), resulting in crossed or uncrossed disparities of the intensity profiles (Fig. 1a,b).

Background luminance was identical to the average stimulus luminance. Between the two stereograms, a dark fixation spot was shown.

3.2 Procedure

The two stereograms S_a and S_a were displayed simultaneously side by side on the computer screen (Fig. 2a,b). In the center of the 2 deg gap between the stereograms, a fixation target was presented that remained visible throughout

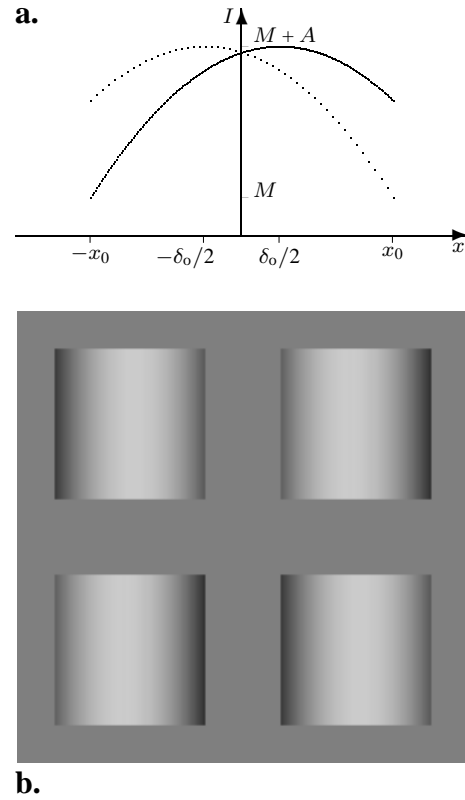


Fig. 1a,b. Stimulus for experiment 1. **a** Intensity profiles without edges (Laplacian zero-crossings) are derived from parabolic arcs. *Continuous line*, $I(x)$; *dotted line*, $I_a(x)$. **b** Example stimulus (schematic). Two stereograms are presented in the upper and lower part of the figure (instead of the left and the right side of the computer screen). The bottom stereogram is a reversed (pseudoscopic) version of the top one. For crossed fusion, the bottom stereogram (S_a) appears in front while the top stereogram (S_a) appears at the back. Note that disparity is constant over all horizontal positions. Perceived curvatures are due to shape-from-shading

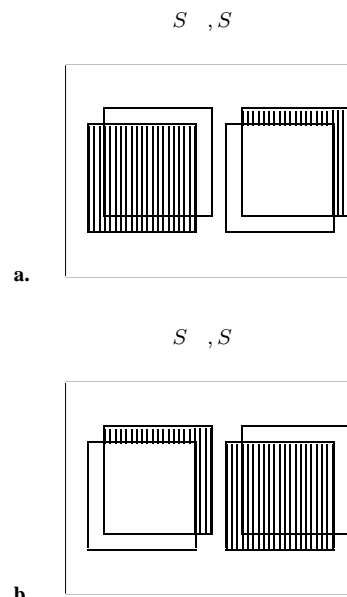


Fig. 2a,b. Procedure for experiment 1. *Hatched squares*, half-image I_a ; *open squares*, half-image I . Subjects had to judge which stereograms were in front

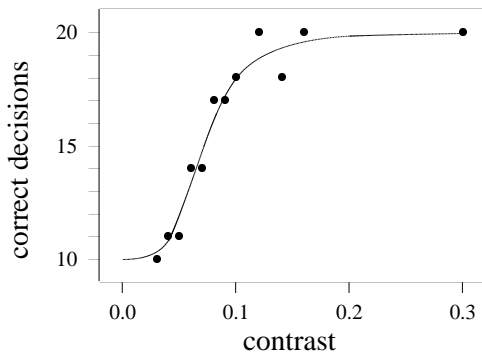


Fig. 3. Sample psychometric function for subject D.J. and disparity 13.2 min arc. *Dots* show the number of correct decisions out of 20 trials. The *curve* is a logistic function fitted to the data using a maximum likelihood criterion. Threshold (75% correct) was reached at contrast level 0.071

the experiment. In a two-alternative forced choice (2AFC) paradigm, subjects had to decide which stereogram (left or right) appeared in front of the other by pressing the left or right button of the computer mouse (depth ordering).

Two-dimensional psychometric functions for disparity (shift, δ_o) and stimulus contrast were measured using the method of constant stimuli at some 70 parameter settings. For each parameter setting, a block of 20 trials was carried out in which the left/right ordering of the two stereograms (S_a, S_a) was varied in pseudorandom fashion.

In the first session, between 10 and 40 trials were run as training. Subjects themselves decided when to begin the measurements. Presentation time for a single trial was not limited. When subjects indicated their decision by pressing a mouse button, the next stimulus was presented after a break of 2 s. Subjects did not get feedback concerning the correctness of their decisions. During the 2 s break, the fixation target remained visible. Subjects were instructed to fixate the target during both the break and stimulus presentation. Eye movements were not recorded. After each block of 20 trials, the stimulus condition (contrast, disparity) was changed. Up to eight blocks of 20 trials were performed in one session. The session ended after 45 min or when the subject reported fatigue or lack of concentration.

From the results for each disparity value (usually about nine contrast values), one-dimensional psychometric functions were calculated using a maximum likelihood fit and a logistic function of the form

$$p(c) = \frac{1}{2} \left(1 + \frac{1 - \epsilon}{1 + \vartheta/c^{1/\sigma}} \right) \quad (8)$$

where c is stimulus contrast, ϑ and σ control threshold and slope, and ϵ is a 'lapse'-factor allowing for erroneous button hits. In our fits, ϵ was chosen to be 0.001. The curves take the value 0.75 at $c = \vartheta$. These values were therefore taken as contrast thresholds. A sample psychometric function is shown in Fig. 3.

3.3 Results

Thresholds are plotted in Fig. 4. Contrast thresholds were lowest at disparities of 15–45 min of arc. At higher dispar-

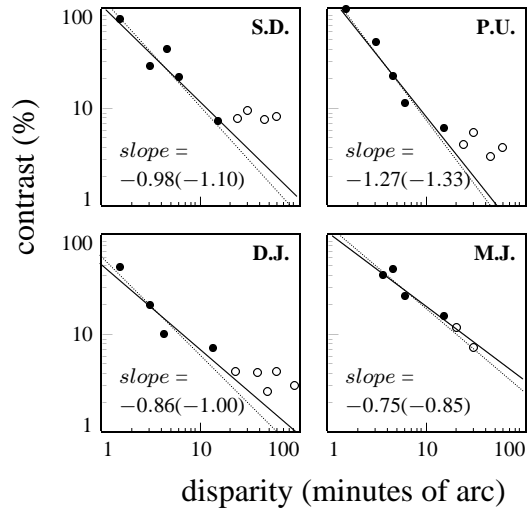


Fig. 4. Contrast thresholds for intensity-based stereo with parabolic luminance profiles (Fig. 1a,b) for four subjects. The *straight lines* are regression curves computed from the data points marked by *filled circles* (disparity $\leq 15'$). The data points marked by *open circles* have been excluded from the computation of the regression lines. *Continuous lines*, contrast as a function of disparity; *dotted lines*, disparity as a function of contrast. The slopes of the dotted lines are given in parentheses

ities, contrast thresholds increased and double images were reported. The asymptote for small disparities is best estimated from data points for disparities of not more than 15 min of arc. Data points above 15 min of arc have therefore not been used to calculate the regression lines in Fig. 4 (open circles). The slope of the regression lines for small disparities is between -0.75 and -1.27 (mean -0.97) for the regression of contrast as a function of disparity (continuous regression lines in Fig. 4). For easier comparison with the data of Legge and Gu (1989), who plotted disparity over contrast, the reciprocals of the slopes of the reverse regression lines should be considered. They range between -0.75 and -1.17 (mean -0.96), i.e., well above the values (around -2) found by Legge and Gu.

3.4 Discussion

The results show that for the parabolic gray wedges used here, disparity threshold is in fact a function of image contrast. The relevant parameter governing the detectability of intensity-based stereo is given by $\text{contrast} \times \text{disparity}^\alpha = \text{const.}$, where α is a subject-specific constant ranging between 0.7 and 1.3 (Fig. 4). This is well in line with the prediction $\alpha = 1.0$ derived for the mean square difference mechanism in Sect. 1.

In a similar threshold measurement with sinusoidal gratings [0.5 and 2.5 cycles per degree (cpd)], Legge and Gu (1989) found an inverse square relationship between contrast and disparity, $c \propto d^{-2}$. From a signal-to-noise analysis, these authors conclude that peak matching rather than matching of zero-crossings or centroids was the most likely stereo mechanism involved. The discrepancy between the results of Legge and Gu and our own measurements may be due to the fact that Legge and Gu's stimulus contained some seven cycles of a sinusoidal grating featuring 14 intensity extrema

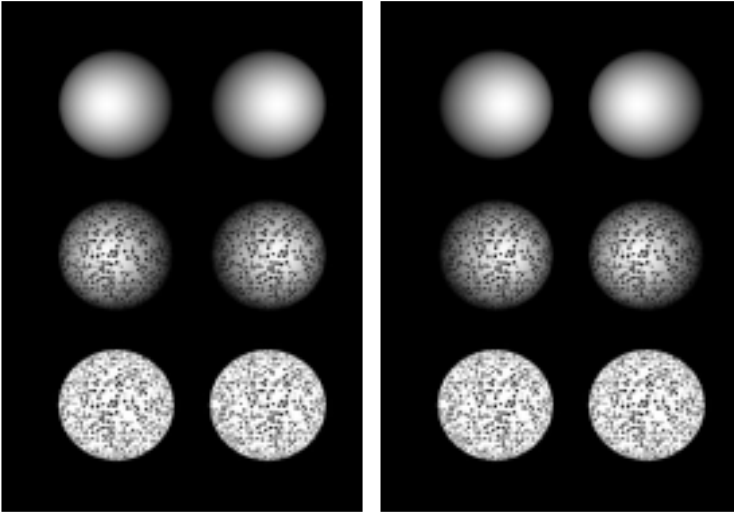


Fig. 5. Demonstration of the depth shift effect associated with intensity-based stereo. The figure shows ellipsoids of revolution. The elongation perpendicular to the paper plane is twice the diameter (elongation $2e_o$). During crossed fusion, orthoscopic versions appear in the left column, and conversely pseudoscopic versions appear in the right column. In uncrossed fusion, the two columns are exchanged. *Top:* In the smoothly shaded ellipsoid, exchange of the half-images results in a perceived overall depth shift while both the orthoscopic and the pseudoscopic version appear convex. *Middle:* If disparate shading and texture information are available, the depth shift effect is still clearly visible. *Bottom:* Only for pure feature-based stereo does exchange of the half-images result in the expected mirroring of the three-dimensional shape at the fixation plane. In the experiments reported in this paper, only the smoothly shaded ellipsoid (top row) is investigated further

and 14 zero-crossings, while our stimulus had only one extremum and no zero-crossings (fundamental frequency 0.08 cpd). The inverse square relationship described by Legge and Gu therefore seems to characterize a feature-matching mechanism, whereas in purely intensity-based stereo mechanisms inverse proportionality is observed.

4 Experiment 2. Quantitative depth from intensity-based stereo

4.1 Stimuli

Simulated surfaces showed the outside (convex) and inside (concave) of ellipsoids of revolution of varying elongation, the axis of revolution being perpendicular to the display screen (Fig. 5, top row). Illumination was simulated from behind the observer and surface reflectance was modelled by the Lambertian cosine law; mutual illumination in the concave object was neglected. Since parallel projection was used, the stereograms of the convex and the concave surface are pseudoscopic versions of each other, i.e., can be generated by exchanging the half-images (cf. Fig. 5). (For a detailed discussion of the geometry of pseudoscopic presentation, see Appendix B.) The stereograms were displayed on a dark background.

The diameter of the ellipsoids in the display plane was 12 cm, corresponding to a visual angle of 6 deg. We denote the basic radius of 6 cm as e_o (for *elongation*). Ellipsoids with four different elongations perpendicular to the monitor surface (3, 6, 12 and 24 cm or 0.5, 1, 2 and 4 e_o) were used in the experiment.

4.2 Procedure

An interactive adjustment task was used to quantify the overall depth separation between orthoscopic and pseudoscopic stereograms ('eggs' and 'bowls': see Fig. 6). The pseudoscopic stereogram was displayed on the left half of the monitor screen with a depth offset z_{pseudo} between e_o corresponding to fixed disparities of 9, 3.6, or 0 min of arc in

pseudorandom order. On the right half of the screen the orthoscopic version of the stereogram appeared in a randomly chosen depth position. Horizontal separation between the two stereograms was 5 cm or 2.5 deg of visual angle. Subjects were asked to adjust the depth position of the orthoscopic stereogram by shifting the half-images horizontally (under mouse control) until it appeared to be at the same depth as the pseudoscopic stereogram. The adjusted depth is denoted by z_{ortho} .

4.3 Results

The pseudoscopic image of a smoothly shaded ellipsoid is perceived as a solid, albeit somewhat flatter object at an increased distance as compared with the orthoscopic presentation. The difference in perceived overall distance increases for more elongated objects, i.e., objects with a greater depth variation (Fig. 7). All subjects reported some uncertainty about how to adjust the orthoscopic stereogram at high elongations of the ellipsoid because the pseudoscopic stereogram appeared much flatter than the orthoscopic one. Subjects were instructed to develop their own strategy to adjust the depth of the probe surface and to use this strategy throughout the experiment. The differences in these strategies are reflected in our results: subject M.J. adjusted depth positions with respect to the tip of ellipsoid, while subject U.B. adjusted the depth of the occluding contours of the ellipsoid.

The depth shift between the egg and the bowl is independent of the absolute position of the fixed object, z_{pseudo} . In Fig. 8 we plotted the relative differences $(z_{\text{ortho}} - z_{\text{pseudo}})/e_o$ as a function of the elongation of the ellipsoid, e/e_o . In a double-logarithmic plot, the resulting data points fall along straight lines with slopes ranging from 1.51 to 2.07.

4.4 Discussion

The results of experiment 2 indicate that the depth averaging performed by intensity-based stereo is contrast dependent. If, in Fig. 6c, the elongation of the ellipsoid is varied by some

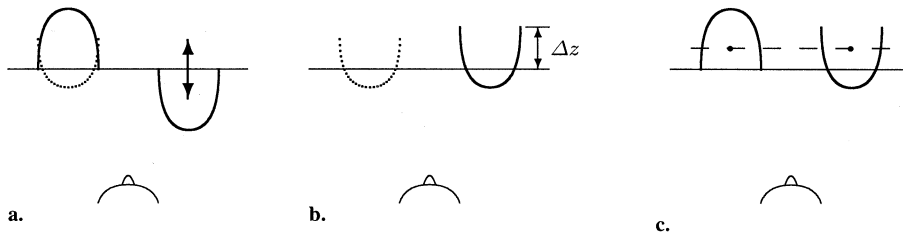


Fig. 6a–c. Bird's eye view of the design of experiment 2. *Continuous lines*, simulated surfaces, *dotted lines*, perceived surfaces. **a** Two stereograms are displayed simulating a convex ellipsoid (*right*) and a concave ellipsoid (*left*). As is demonstrated in Fig. 5, the shaded concave version will look like a less elongated convex ellipsoid, somewhat further away from the observer. The subject can interactively shift the orthoscopic ellipsoid in depth. **b** In an adjustment task, the perceived distances of the two surfaces are equated. **c** A possible interpretation is that the average depth of both surfaces (indicated by the *black dot*) is equated

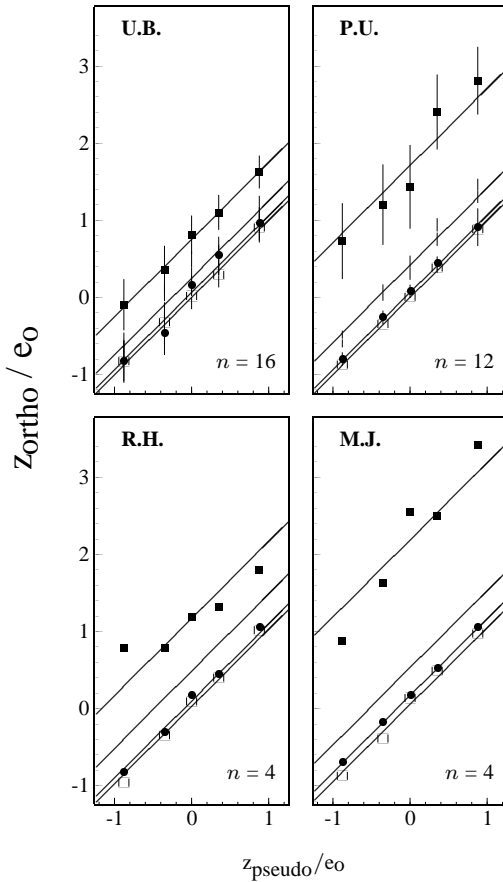


Fig. 7. Results of experiment 2. z_{ortho} , adjusted depth position of the orthoscopically presented ellipsoid; z_{pseudo} , depth position of the pseudoscopic version. All distances are given in multiples of $e_o = 6\text{ cm}$. n , number of adjustments per data point. Relative elongations of the ellipsoids were: *open squares*, 0.5; *filled circles*, 1.0 (i.e., a sphere); *open circles*, 2.0 (i.e., an egg); *filled squares*, 4.0. The mean depth offset for each elongation is indicated by the oblique lines (slope 1.0)

multiplicative factor, the depth average symbolized by the black dot does not move by the same factor. This finding can be predicted from the square difference mechanism (6). This equation stated that global perceived depth corresponds to a contrast-weighted average of the local veridical disparities.

To calculate the predicted depth offset from the square difference model, we denote the depth profile of an ellipsoid with elongation e by $z_e(r, \varphi) := e\sqrt{1-r^2}$, $r \in [0, 1]$ and note that in parallel projection, z is proportional to (relative)

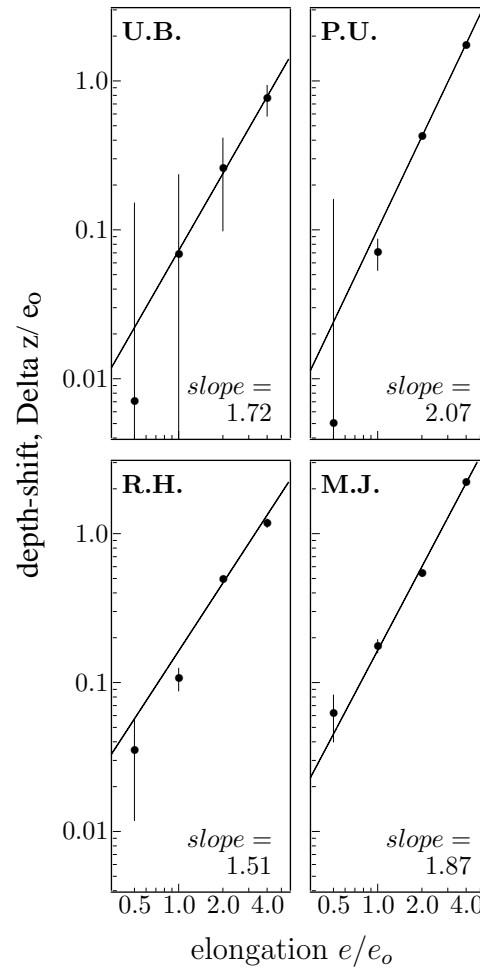


Fig. 8. Relative perceived offset between the ortho- and pseudoscopic view of an ellipsoid plotted as a function of elongation. The data are replotted from Fig. 7 by averaging the relative depth offset over the five absolute depth positions presented in experiment 2. *Error bars* are standard deviations of the mean. The long bars, in particular those for subjects U.B. and P.U., are a result of logarithmic scaling of small values. The *straight lines* are regression lines minimizing the error-weighted square deviations. The slopes range from 1.51 to 2.07, i.e., well above the value of 1.0 predicted from contrast-independent averaging

disparity. From the profile z we can calculate the image intensity distribution $I_e(r, \varphi)$ for frontal illumination by means of Lambert's cosine law (cf. Bülthoff and Mallot 1988):

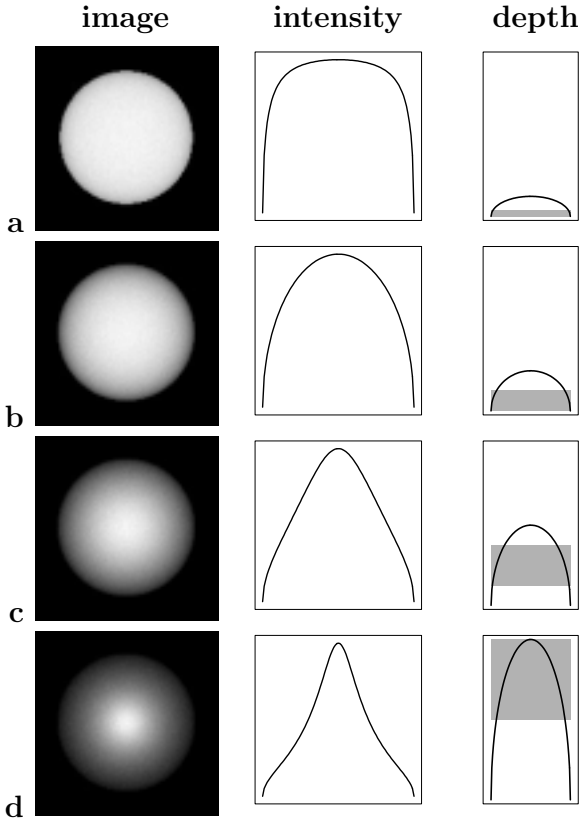


Fig. 9a–d. Contrast-weighted depth averaging in experiment 2 [cf. (11)]. *Left-hand column*, Gray-level image. *Middle column*, intensity profile [section through gray-level image, cf. (9)]; *x*-axis, image position (ranges from -1 to 1); *y*-axis, intensity (ranges from 0 to 1). *Right-hand column*, depth profile; *x*-axis, image position (ranges from -1 to 1); *y*-axis, depth (ranges from 0 to 4). **a–d** Elongations 0.5 , 1.0 , 2.0 , and 4.0 , respectively. The shaded areas in the depth profiles indicate regions with high image contrast. In contrast-weighted depth averaging, these regions will receive higher weights

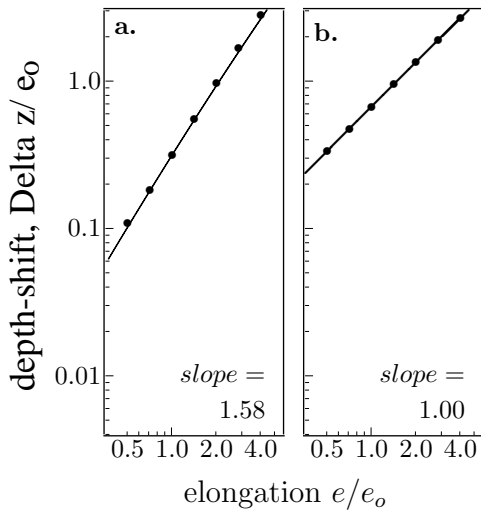


Fig. 10a,b. Predictions for the perceived depth offset shown in Fig. 8. **a** Contrast-weighted averaging as predicted by the mean square difference mechanism (1, 11) results in a straight line with a slope of 1.58 . **b** Contrast-independent weights predict a slope of 1.00

$$I_e(r, \varphi) = \left(\frac{1 - r^2}{1 - (1 - e^2)r^2} \right)^{\frac{1}{2}} \quad (9)$$

The weight function $w(r, \varphi)$, i.e., the square of the partial derivative of I with respect to the cartesian image coordinate x [horizontal direction: see (6)] is given by

$$\begin{aligned} w_e(r, \varphi) &= \cos^2 \varphi \left(\frac{\partial I_e(r, \varphi)}{\partial r} \right)^2 \\ &= \frac{(er \cos \varphi)^2}{(1 - r^2)(1 - r^2 + e^2 r^2)^3} \end{aligned} \quad (10)$$

Substituting these results into (6) yields the prediction:

$$\Delta z(e) = \frac{\int_{-} \int_0^1 z_e(r, \varphi) w_e(r, \varphi) r dr d\varphi}{\int_{-} \int_0^1 w_e(r, \varphi) r dr d\varphi} \quad (11)$$

The weight function $w_e(r, \varphi)$ has a singularity of order two for $r = 1$, i.e., at the occluding contour of the imaged ellipsoid. Thus, strictly speaking, the integrals in (11) do not converge. However, this problem arises only if infinite image resolution is assumed. To account for finite resolution, we limit the integration to the inner pixels of the display (i.e., $r \in [0, 0.995]$ rather than $[0, 1]$).

The argument is summarized in Fig. 9a–d. As the elongation of the imaged ellipsoid increases, the regions of high image contrast move from the outline towards the center of the ellipsoid’s image. At low elongations, the average is thus dominated by marginal parts of the image, i.e., regions with relatively little depth, whereas at high elongations, the average is taken predominantly from regions with high depth values. As a result, the depth average grows faster than the elongation.

The numeric solution of (11) can be approximated by a power law with exponent 1.58 . In a double-logarithmic plot, the theoretical curve thus becomes a straight line with a slope of 1.58 (Fig. 10a). This prediction is well in line with the measured slopes, which range from 1.51 to 2.07 .

As an alternative possibility, assume that the weighting function does not depend on local contrast. In this case, it is independent of the ellipsoid’s elongation as well, since contrast distribution is the only stimulus parameter influenced by elongation in our experiments. A purely spatial weighting function $w(r)$ might still account for differences in saliency of central and peripheral regions of the ellipsoid’s image. From (11), we can analytically compute the relation of Δz and e for contrast-independent weighting by using the relation $z_e(r, \varphi) = e z_1(r, \varphi)$:

$$\Delta z(e) = e \frac{\int_{-} \int_0^1 z_1(r, \varphi) w(r) r dr d\varphi}{\int_{-} \int_0^1 w(r) r dr d\varphi} = k e \quad (12)$$

for some constant k depending on the spatial weight distribution. For contrast-independent weighting, (12) predicts a linear relationship between elongation and perceived offset; the predicted slope of the regression lines in the double-logarithmic plot of Fig. 8 is therefore 1.0 (Fig. 10b). Note that the case of contrast-independent weighting also includes the possible use of the point-disparity of the intensity extremum (which here does provide some disparity information) or of the shallow Laplacian zero-crossings occurring

for $e > 1$; in these cases, perceived offset should again be proportional to elongation.

In summary, our data strongly suggest that the weighting function does depend on contrast, since any contrast-independent weighting would result in a slope of 1.0. As a possible candidate for weighting, we propose the square of the image intensity gradient (or local power) as a measure of local contrast. This weighting function explains the results of experiment 2 and is predicted by the mean square difference mechanism of intensity-based stereo.

5 General discussion: Modularity of stereoscopic depth perception

5.1 Feature-based versus intensity-based stereopsis

The data presented here and in the previous paper (Arndt et al. 1995) strongly suggest the existence of a mean square difference mechanism in human stereopsis. It should be understood that the distinction between mean square difference, which has to be minimized in order to obtain disparity, and interocular correlation, which has to be maximized, is subtle, lying mostly in the normalizations of the intensity function involved. Since normalization was not addressed in our experiments, we use the two terms synonymously in this discussion.

Our result is well in line with a number of earlier findings including, among others, global surface perception in random dot stereograms superimposed with uncorrelated noise (Cormack et al. 1991), the perception of ghost planes in random arrays of double-nail displays (Weinshall 1991), and the computation of average depth for disparity-evoked vergence (Mallot et al. 1996b). Interestingly, all these examples of a mean square difference mechanism relate to surface perception or global stereopsis in the sense of Julesz (e.g., Julesz 1971).

Do these findings justify the distinction between different independent mechanisms or ‘modules’ in stereo processing? More specifically, can we separate a correlation module from a feature correspondence module? One argument in favor of this distinction is that a (hypothetical) module relying on feature correspondence alone could not ‘see’ intensity-based stereo while a (likewise hypothetical) module relying exclusively on mean square difference would fail to localize single dots in three-dimensional clouds. However, it is easy to conceive of stereo mechanisms with adjustable smoothness or correlation windows which approximate global correlation if only coarse image data are available while higher resolutions are obtained in the presence of sufficient gray-level variation (i.e., features; for an example, see DeAngelis et al. 1995). Bülthoff and Yuille (1991) have formulated the idea of stimulus-dependent smoothing in the framework of the Bayesian theory of estimation. Experimental support for this idea comes from an experiment reported by Bülthoff et al. (1991) who showed that the dependence of perceived depth on disparity gradient is modulated by stimulus properties such as the shape of the matching targets. Neural network realizations of Bayesian algorithms are easy to construct using disparity-tuned units (for review see Blake and Wilson 1991) which would be capable of evaluating both smooth and feature-based input information.

An empirical argument supporting the separation of two modules comes from the study of the role of contrast in images with different amounts of edge-based information. If good edge information is available, contrast does not influence stereo acuity as long as the stimulus remains visible (Lit et al. 1972). However, if edge information is blurred, such as in gratings with low spatial frequency, contrast becomes important. For example, Legge and Gu (1989) have shown that for gratings with 0.5 and 2.5 cycles per degree, contrast at stereo threshold is inversely proportional to disparity squared. From this result, they concluded that a peak-matching mechanism is involved. This conclusion was based on a consideration of the positioning error of intensity peaks as a function of the signal-to-noise ratio. Since we have shown that the mean square difference mechanism predicts an inverse proportionality between contrast and disparity (5), the difference between our result for images without edges (slope -1 in Fig. 4) and the result of Legge and Gu (1989) for images with blurred edge and peak information, hints at different mechanisms for feature- and intensity-based stereopsis. It should be stressed, however, that even if a distinct module for feature-based stereo exists, it is not entirely free of contrast information: contrast similarities do influence the selection of stereo matches (Smallman and McKee 1995; Mallot et al. 1996a).

A dissociation between intensity- and feature-based image data was reported by Arndt et al. (1995, Sect. 1.3): in a pseudoscopic presentation of dotted ellipsoids (as in the middle row of Fig. 5), some subjects perceived veridical depth from the dots (i.e., a concavity) while intensity information led to the perception of a transparent convex ‘dome’ in front of the dots. It is not entirely clear, however, whether this result is due to a dissociation between intensity- and feature-based stereopsis, or between shape-from-shading and feature-based stereopsis.

5.2 Intensity-based stereopsis versus shape-from-shading

In contrast to feature-based stereopsis, intensity-based stereopsis does not lead to a pointwise depth map and thus no shapes can be derived. For example, a pseudoscopic stereogram of a textured ellipsoid (i.e., a stereogram of a bowl) is seen as a concave surface. In contrast, a pseudoscopic stereogram of a smooth shaded ellipsoid is not seen concave but as a solid object which appears flatter and further away than its orthoscopic version (cf. Fig. 5). Thus, intensity-based stereo determines the perceived distance of the entire object but cannot overcome pointwise data from monocular cues such as shape-from-shading or prior assumptions such as convexity and general viewpoint.

5.3 An ecological view of modularity in depth perception

In summary, we propose the following preliminary framework for further investigation of stereo and depth mechanisms (cf. Fig. 11). We start from an ecological perspective of depth perception as a process allowing recovery of various descriptions of three-dimensional objects (shape, distance, depth ordering, orientation, etc.) from depth cues in

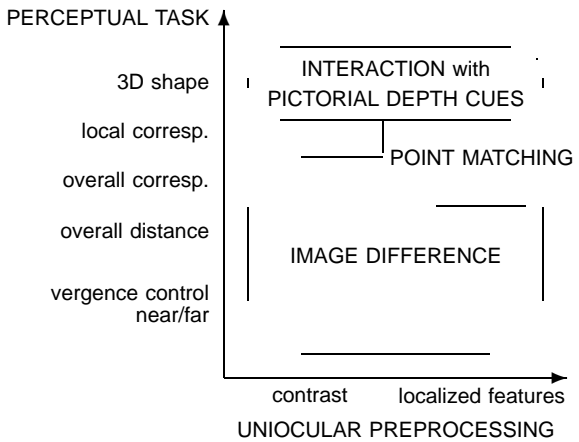


Fig. 11. Proposed breakdown of stereo mechanisms requiring different amounts of unocular preprocessing and subserving different perceptual competences

the visual sensory input (disparities, shading, texture gradients, motion parallax, to name just a few). In this view, modules are behavioral, not computational subunits. We use the term ‘depth descriptor’ to denote different experimentally assessable aspects of depth perception without implying that for each ‘descriptor’ there is a specialized computational module, representation or percept (Bülthoff and Mallot 1990; see also Braunstein et al. 1986). We can now characterize putative stereo mechanisms along two dimensions:

A. Which type of image information is used and how much unocular preprocessing is required? The simplest cases here are shape-from-shading and intensity-based stereo working directly on image contrast. Successively higher stages of preprocessing are required for feature-based point disparities, texture gradients, orientation and motion disparities, etc.

B. Which depth descriptor is the mechanism used for? Here, we propose the following stages of complexity: 1. *Depth ordering or near/far distinction* is required for orientation reactions including vergence control. In Arndt et al. (1995) as well as in experiment 1 of the present paper, we have shown that intensity-based disparities suffice to perform this task. In vergence control it was shown that even if edge information is present, contrast information is used in a way predicted quantitatively by a correlation scheme (Cormack et al. 1991; Mallot et al. 1996b).

2. *Overall distance of objects* is a quantitative version of the global depth ordering task mentioned above. In experiment 2, we showed that this task again can be solved purely from intensity-based disparity information.

3. *Overall correspondence* is studied, e.g., with the wall-paper illusion (see McKee and Mitchison 1988; Jordan et al. 1990). This is the simplest case in which the problems of stereo correspondence arise, since only one global disparity has to be found for the entire image. Mallot et al. (1996a) have shown that contrast information is involved in this task as well. If conflicting information is presented in different spatial frequency bands, the distribution of signal energy (or autocorrelation) across the different channels determines depth perception (see also Smallman 1995).

4. *Correspondence of localized image features* yields sparse depth information at more or less isolated points. This is not easily explained by a mean square difference mechanism. As mentioned above, contrast information seems not to be involved in this task (but see Smallman and McKee 1995). Other characteristics of correlation schemes, such as distance weighting (decreased plausibility of high disparity matches), however, have been found as well (Mallot and Bideau 1990).

5. *Three-dimensional shape* is the richest and most complex description of an imaged surface. In human shape perception, pictorial cues such as shading or texture gradients as well as prior assumptions play an important role (Bülthoff and Mallot 1988; Buckley and Frisby 1993).

Figure 11 summarizes our view of different mechanisms involved in depth perception. The simplest (and possibly evolutionarily oldest) mechanism is mean square difference or correlation using raw image information and leading to simple, coarse stereoscopic percepts. This mechanism is strongly affected by image contrast and produces a coarse description of continuous surfaces in space. Its mechanism is computationally characterized by area correlation or the mean square difference of image intensities [cf. (1)]. Besides intensity-based stereo, we propose including global stereopsis, i.e., the perception of continuous surfaces based on large numbers of (random) dots (Julesz 1971) in this module. This is supported by the fact that in experimental procedures favoring the perception of global stereopsis, correlation-type mechanisms have been found similar to the square difference model proposed here for intensity-based stereo. The ecological usefulness of image difference stereopsis has been demonstrated in, for example, binocularly guided prey catching (Wiggers et al. 1995) and obstacle avoidance by autonomous vehicles (Mallot et al. 1990).

In contrast, the correspondence mechanism or standard feature-based stereopsis operates in the presence of (not too many) localized image features. It is only weakly affected by image contrast (except for contrast polarity) and leads to the perception of isolated points in space. Observations including autostereograms and the double-nail illusion (Krol and van de Grind 1980) indicate that feature-based stereo actually suffers from the correspondence problem and relies on ordering constraints of various kinds to overcome it. If crossed with intensity-based stereo, either it vetos the contradictory information, or surfaces and dots corresponding to the two conflicting cues are perceived simultaneously at different depth positions (‘subjective surfaces’: Bülthoff and Mallot 1988; Arndt et al. 1995).

6 Conclusion

We have shown that minimization of mean squared image difference can account quantitatively for stereoscopic depth perceptions evoked by smoothly shaded images. The contrast and disparity thresholds for intensity-based stereo are inversely proportional to each other, as is predicted by the square difference mechanism. The amount of depth perceived from curved surfaces without localized features is an

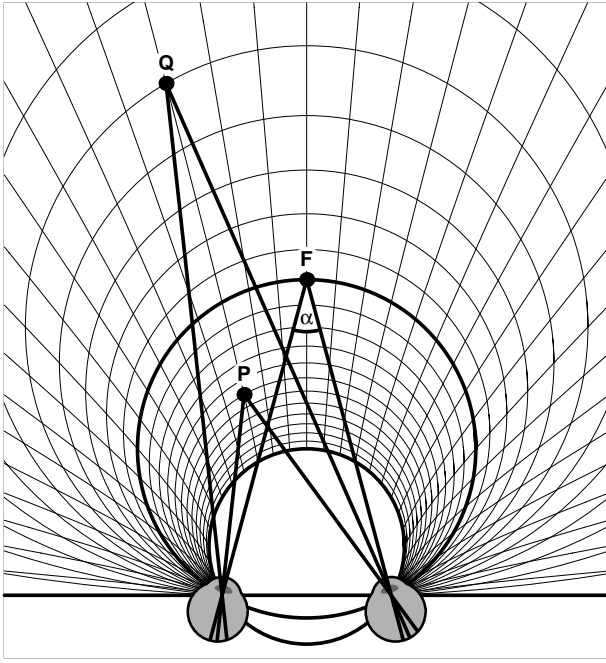


Fig. B1. Geometry of pseudoscopic image presentation. F , fixation point; the angle α at F is the vergence angle. *Circles* are iso-disparity curves (spacing 2.5°), including the Vieth–Müller circle passing through fixation. *Radial lines* are the outer parts of the hyperbolae of Hillebrand, that is iso-curves of cyclopean direction or Hering version, γ , i.e., lines where the mean of the angles with the left and right lines of sight is constant (spacing: 5°). The *open area* enclosed by the iso-disparity circle with disparity $-\alpha$ is not mapped to a physically meaningful position by pseudoscopic inversion

average of the true disparities weighted with the local image contrast. Again, this result was predicted by the square difference mechanism.

Appendix A. Derivation of Eq. 6

We have to minimize the expression (1)

$$\Phi(D) = \int \int [I_l(x, y) - I_r(x + D, y)]^2 dx dy \quad (\text{A1})$$

Let us denote by $\delta(x, y)$ the true disparity profile, i.e., a function describing the simulated surface. For the sake of simplicity, we use the asymmetric formulation:

$$I_r(x, y) = I_l(x - \delta(x, y), y) \quad (\text{A2})$$

Setting $I(x, y) := I_l(x, y)$ and substituting (A2) into (A1), we obtain:

$$\Phi(D) = \int \int [I(x, y) - I(x - \delta(x, y) + D, y)]^2 dx dy \quad (\text{A3})$$

The minimization is now performed by setting $\Phi'(D) = 0$. The solution will be denoted by D . Applying the chain rule,

$$0 = 2 \int \int I(x - \delta(x, y) + D, y) \times [I(x, y) - I(x - \delta(x, y) + D, y)] dx dy \quad (\text{A4})$$

and linearly approximating the term in brackets,

$$0 = \int \int I(x, y) [(D - \delta(x, y)) I(x, y)] dx dy \quad (\text{A5})$$

we finally obtain:

$$D = \frac{\int \int \delta(x, y) I^2(x, y) dx dy}{\int \int I^2(x, y) dx dy} \quad (\text{A6})$$

Appendix B. Geometry of pseudoscopic stereograms

In the main section of the paper, we assumed that exchanging the half-images of a stereogram simulates a surface which is roughly mirrored in depth at the zero-disparity plane. This simplifying assumption is not completely correct, as will be discussed in this appendix. The deviations are only minor, however, and do not affect the argument presented above.

The geometry of the pseudoscopic inversion of single points is summarized in Fig. B1. Consider a point P at angles φ_l and φ_r measured from the left and right axes of sight, respectively. Its disparity (relative to the fixation point) is given by $\delta := \varphi_l - \varphi_r$; its cyclopean direction is $\gamma := (\varphi_l + \varphi_r)/2$. With these notations, the exchange of the half-images amounts to the transformation (pseudoscopic inversion)

$$\mathcal{P} : (\delta, \gamma) \rightarrow (-\delta, \gamma) \quad (\text{B1})$$

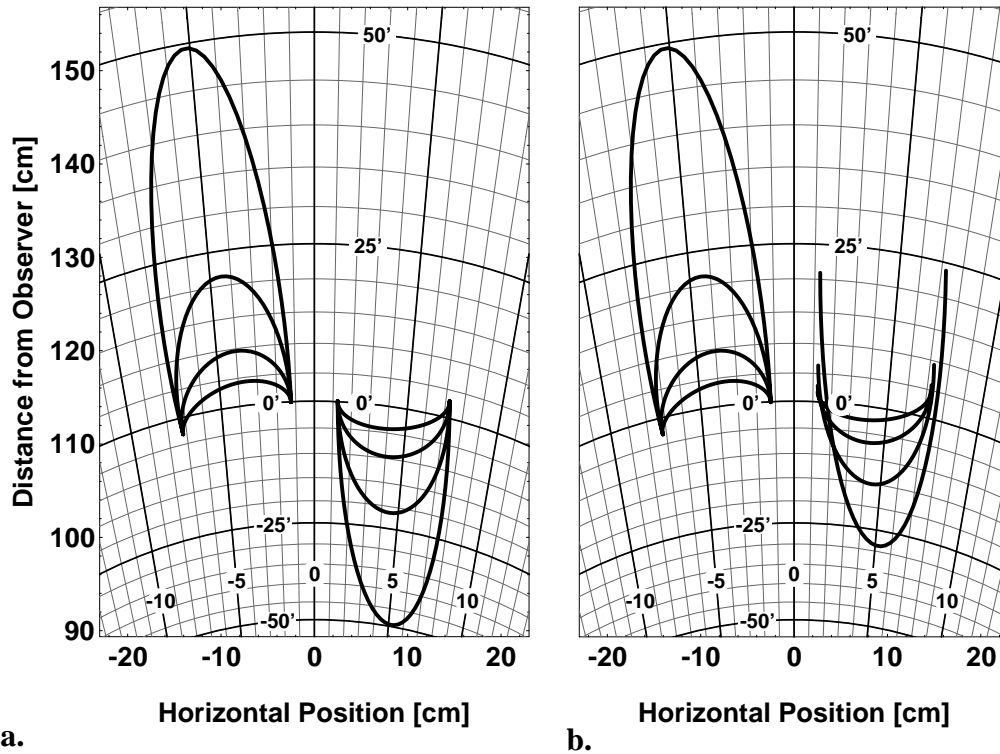
In Fig. B1, the points P and Q are related by pseudoscopic inversion.

Figure B2 shows the pseudoscopic inversion for the surfaces and experimental setup used in experiment 2. In summary, the following deviations from the simple mirroring assumption can be inferred from the two figures:

1. Points with disparities less than (more negative than) $-\alpha$ (the vergence angle of the system) cannot be inverted. The iso-disparity circle for $\delta = -\alpha$ is shown as a bold line in Fig. B1. It is transformed into a ‘circle’ with infinite radius, the proximal part of which is shown as a horizontal line passing through the projection centers. Therefore, pseudoscopic inversion does not lead to geometrically meaningful stereograms in systems with parallel camera axes ($\alpha = 0$) or with small vergence angles; in this case, the rays from the exchanged image points through their corresponding projection centers will diverge. Parallel axes of view occur, for example, when viewing large outdoor scenes such as landscapes. As can be expected from Fig. B1, pseudoscopic presentation of such outdoor scenes does not lead to depth inversions.

2. Even if the point falls in the transformable disparity range ($\delta \in [-\alpha, \alpha]$), the surface it belongs to will be distorted (Fig. B2a). While a true mirroring with respect to a circle is conformal, the pseudoscopic inversion is not.

3. In complex scenes, some points may be occluded for one eye but not for the other. This monocular occlusion (sometimes also called ‘DaVinci stereopsis’), which represents a strong stereoscopic depth cue in itself (see Anderson and Nakayama 1994), is confused by pseudoscopic inversion: the occluding object will now appear further away and the formerly occluded points should be visible from both eyes. The effect can be observed in the bottom row of Fig. 5, where it makes it hard to perceive the pseudoscopic version at all (cf. Braunstein et al. 1986).



a. **b.**
Fig. B2. **a** Simulated depth profiles in the pseudoscopic (*left*) and orthoscopic (*right*) viewing conditions for ellipsoids with four elongations (0.5, 1.0, 2.0, 4.0). **b** Average adjusted depth positions of the orthoscopically presented surfaces for subject M.J. (replotted from Fig. 8). Spacing of iso-disparity-circles: 5 min arc. Spacing of the iso-lines of cyclopean direction: 1°. In the calculation, it is assumed that the subject did not perform vergence eye movements during the adjustment. This was not controlled in the experiments and vergence movements are in fact likely to occur

4. Finally, pseudoscopic presentation of a convex surface results in a concave curvature. If shading is included, this concave surface lacks the interreflection from the inward-slanted walls expected for a realistic image. Thus, shading and stereo information will usually be in conflict in pseudoscopic images.

Acknowledgements Parts of this work were supported by the Deutsche Forschungsgemeinschaft. We thank Sabine Gillner and Anke Roll for helpful discussions, Pia Unrath for help with experiment 2, and Guy M. Wallis for correcting the English text.

References

- Anderson BL, Nakayama K (1994) Toward a general theory of stereopsis: binocular matching, occluding contours, and fusion. *Psychol Rev* 101:414 – 445
- Arndt PA, Mallot HA, Bühlhoff HH (1995) Human stereovision without localized image-features. *Biol Cybern* 72:279 – 293
- Bishop PO, Henry GH, Smith CJ (1971) Binocular interaction fields of single units in the cat striate cortex. *J Physiol (Lond)* 216:39 – 68
- Blake R, Cormack RH (1979) Does contrast disparity alone generate stereopsis? *Vision Res* 19:913 – 915
- Blake R, Wilson HR (1991) Neural models of stereoscopic vision. *Trends Neurosci* 14:445 – 452
- Braunstein ML, Anderson GJ, Rouse MW, Tittle JS (1986) Recovering viewer-centered depth from disparity, occlusion, and velocity gradients. *Percept Psychophys* 40:216 – 224
- Buckley D, Frisby JP (1993) Integration of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Res* 33:919 – 933
- Bühlhoff HH, Fahle M, Wegmann M (1991) Perceived depth scales with disparity gradient. *Perception* 20:145 – 153

- Bühlhoff HH, Mallot HA (1988) Integration of depth modules: stereo and shading. *J Opt Soc Am A* 5:1749 – 1758
- Bühlhoff HH, Mallot HA (1990) Integration of stereo, shading and texture. In: Blake A, Troscianko T (eds) *AI and the eye*. Wiley, Chichester, pp 119 – 146
- Bühlhoff HH, Yuille A (1991) Bayesian models for seeing shapes and depth. *Comments Theor Biol* 2:283 – 314
- Christou CG, Parker AJ (1993) An investigation of intensity-based stereo in human shape judgements. *Perception* 22 [Suppl]:106
- Cormack LK, Stevenson SB, Schor CM (1991) Interocular correlation, luminance contrast and cyclopean processing. *Vision Res* 31:2195 – 2207
- DeAngelis GC, Ohzawa I, Freeman RD (1995) Neuronal mechanisms underlying stereopsis: how do simple cells in the visual cortex encode binocular disparity? *Perception* 24:3 – 31
- Foley JD, van Dam A, Feiner SK, Hughes JF (1990) *Computer graphics: principles and practice*, 2nd edn., Addison-Wesley, Reading, Mass
- Hodges LF (1992) Time-multiplexed stereoscopic computer graphics. *IEEE Comput Graphics Applications* 12(2):20 – 30
- Howard IP, Rogers BJ (1995) *Bionocular vision and stereopsis*. (Oxford Psychology Series no. 29) Oxford University Press, Oxford
- Jordan JR III, Geisler WS, Bovik AC (1990) Color as a source of information in the stereo correspondence process. *Vision Res* 30:1955 – 1970
- Julesz B (1971) *Foundations of cyclopean perception*. Chicago University Press, Chicago
- Krol JD, van de Grind WA (1980) The double-nail illusion: experiments on binocular vision with nails, needles, and pins. *Perception* 9:651 – 669
- Legge GE, Gu Y (1989) Stereopsis and contrast. *Vision Res* 29:989 – 1004
- Lehky SR, Sejnowski TJ (1990) Neural model of stereoacuity and depth interpolation based on a distributed representation of stereo disparity. *J Neurosci* 10:2281 – 2299
- Lit A, Finn JP, Vicars W (1972) Effect of target background luminance contrast on binocular depth discrimination at photopic levels of illumination. *Vision Res*, 12:1241 – 1251
- Mallot HA, Bideau H (1990) Binocular vergence influences the assignment of stereo correspondences. *Vision Res* 30:1521 – 1523

- Mallot HA, Gillner S, Arndt PA (1996a) Is correspondence search in human stereo vision a coarse-to-fine process? *Biol Cybern* 74:95 – 106
- Mallot HA, Roll A, Arndt PA (1996b) Disparity-evoked vergence is driven by interocular correlation. *Vision Res*, in press
- Mallot HA, Zielke T, Storjohann K, von Seelen W (1990) Topographic mapping for stereo and motion processing. In: Casasent DP (ed) *Intelligent robots and computer vision IX: Neural biological and 3-D methods* (Proceedings vol. 1382), pp 397 – 408
- Marr D, Poggio T (1976) Cooperative computation of stereo disparity. *Science* 194:283 – 287
- Marr D, Poggio T (1979) A computational theory of human stereo vision. *Proc R Soc (Lond) B* 204:301 – 328
- Mayhew JEW, Frisby JP (1981) Psychophysical and computational studies towards a theory of human stereopsis. *Artif Intell* 17:349 – 385
- McKee SP, Mitchison GJ (1988) The role of retinal correspondance in stereoscopic matching. *Vision Res* 28:1001 – 1012
- Ohzawa I, DeAngelis GC, Freeman RD (1990) Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science* 249:1037 – 1041
- Poggio GF (1995) Mechanisms of stereopsis in monkey visual cortex. *Cerebral Cortex* 5:193 – 204
- Poggio GF, Fischer B (1977) Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkeys. *J Neurophys* 40:1392 – 1407
- Qian N (1994) Computing stereo disparity and motion with known binocular cell properties. *Neural Comput* 6:390 – 404
- Richards WA (1971) Anomalous stereoscopic depth perception. *J Opt Soc Am* 61:410 – 414
- Smallman HS (1995) Fine-to-coarse scale disambiguation in stereopsis. *Vision Res* 35:1047 – 1060
- Smallman HS, McKee SP (1995) A contrast ratio constraint on stereo matching. *Proc R Soc (Lond) B* 260:265 – 271
- Weinshall D (1991) Seeing 'ghost' planes in stereo vision. *Vision Res* 31:1731 – 1748
- Westheimer G, McKee SP (1980) Stereoscopic acuity with defocused and spatially filtered retinal images. *J Opt Soc Am* 70:772 – 778
- Wiggers W, Roth G, Eurich C, Straub A (1995) Binocular depth perception mechanisms in tongue-projecting salamanders. *J Comp Physiol [A]* 176:365 – 377