Perception, 2006, volume 35, pages 145 – 155

DOI:10.1068/p5315

Disparity and shading cues cooperate for surface interpolation

Quoc C Vuong¶§, Fulvio Domini§, Corrado Caudek#
¶ Max Planck Institute for Biological Cybernetics, D 72076 Tübingen, Germany; § Department of Cognitive and Linguistic Sciences, Brown University, Box 1970, Providence, RI 02912-1978, USA; # Department of Psychology, University of Trieste, via S Anastasio 12, I 34100 Trieste, Italy;
e-mail: quoc.vuong@tuebingen.mpg.de
Received 14 April 2004, in revised form 14 March 2005; published online 18 January 2006

Abstract. In two experiments, we tested whether disparity and shading cues cooperated for surface interpolation. Observers adjusted a probe dot to lie on a surface specified either by a sparse disparity field, a continuous stereo shading or monocular shading gradient, or both cues. Observers' adjustments were very consistent with disparity information but their adjustments were much more variable with shading information. However, observers significantly improved their precision when both cues were present, relative to when only disparity information was present. These results cannot be explained by assuming that separate modules analyze disparity and shading information, even if observers optimally combined these cues. Rather, we attribute this improvement to a process through which the shading gradient constrains the disparity field in regions where disparities cannot be directly measured. This cooperative process may be based on the natural covariation existing between these cues produced by the retinal projection of smooth surfaces.

1 Introduction

Properties of objects and surfaces (eg depth, orientation, and shape) are specified by multiple sources of information on the retina, such as binocular disparity, shading, and image velocity, among others. Combined, these cues converge on unified estimates of the different properties, which is necessary both for navigation, and for manipulating and recognizing objects. Separately, however, each cue provides a noisy, and sometimes different, estimate of the *same* distal property. How does the human visual system make sense of these estimates to form coherent percepts? The prevalent answer is that the visual system takes a weighted average of independent estimates, with the weights inversely proportional to the reliability of the cue in the scene (Clark and Yuille 1990; Ernst and Banks 2002; Jacobs 1999; Landy et al 1995).

A strong assumption in the literature is that *independent* modules estimate depth at each point in the scene; that is, the output of each module is a point-wise depth map (Landy et al 1995). However, each cue provides a qualitatively different kind of depth map; for example, occlusions in the scene provide only ordinal depth at each spatial position, whereas motion parallax (in principle) provides absolute (ie metric) depth. Thus, some form of interaction or cooperation between modules is unavoidable if these estimates are to be combined. To constrain the form of this interaction, Landy et al (1995) proposed a modified weak fusion (MWF) model that permits interactions between modules solely for the purpose of *promoting* qualitatively different depth maps to absolute (ie common) units that can then be averaged. How might cue promotion occur? Landy et al suggested that, since isolated cues compute depth maps with a number of missing parameters (eg fixation distance, vergence angle, sign of rotation, and so on), different cues could set values for these missing parameters by, for example, solving a system of equations (eg Johnston et al 1994).

Here we examined the interaction between disparity and shading cues for surface interpolation. Previous work by Lappin and Craft (2000) and by us (Vuong et al 2004) demonstrated that human observers could precisely interpolate surfaces from sparse disparity fields. The degree of precision with which observers can determine depth in regions without disparities is quite impressive because there is, in principle, an infinite number of solutions to the interpolation problem. It should be noted that interpolating a sparse disparity field is not trivially based on comparisons of local disparity signals (Vuong et al 2004). Rather, as proposed by Lappin and Craft (2000), the human visual system may solve the interpolation problem by approximating the sparse disparity field with a smooth parabolic surface. In the present study, we develop this idea further by suggesting that the visual system may use isomorphic relations existing between different depth cues as constraints for surface interpolation.

In an earlier study most similar to ours, Bülthoff and Mallot (1988) examined how observers estimated depth at several simultaneously probed locations of an ellipsoid from either *stereo disparity* (provided by localized edges), *stereo shading* (the weak disparity signal from the shifted luminance gradient in each monocular image; what the authors called disparate shading), *monocular shading* (the luminance gradient in each monocular image), or combinations of these cues. Their main findings were that, first, the amount of perceived depth increased with additional cues; and, second, disparate edges effectively 'vetoed' stereo shading and monocular shading cues (in the case of cue conflict). From these results, the authors hypothesized an accumulation of depth from available cues, with nonlinear interactions between modules (eg vetoing and inhibition). Landy et al (1995) suggested that nonlinear interactions could arise from cue promotion, since, for instance, monocular shading only indicates the local orientation of a shape at a particular location, rather than depth per se at that location. The presence of a strong disparity signal could promote the shading cue to an absolute depth cue leading to the observed interactions.

Given the problem of surface interpolation, however, it is not clear how shading cues could promote stereo cues in regions where there are no disparity signals. However, if we consider that a linear relationship exists between the second derivative of the disparity field and the luminance gradient, then it follows that disparity values can be directly estimated from the smooth shading gradient. This process is different from the promotion stage proposed by Landy et al (1995), since it does not require any direct estimate of 3-D properties. Indeed, for the case of stereo and shading, the following image relation must necessarily be satisfied (Horn 1986):

$$\begin{bmatrix} \dfrac{\partial d_\alpha}{\partial \gamma} \\[2ex] \dfrac{\partial d_\beta}{\partial \gamma} \end{bmatrix} = D \begin{bmatrix} I_x \\ I_y \end{bmatrix}, \tag{1}$$

where $\dfrac{\partial d_\alpha}{\partial \gamma}$ and $\dfrac{\partial d_\beta}{\partial \gamma}$ represent the variation of the horizontal and vertical gradients of the disparity field in the direction of maximum image luminance change; I_x and I_y represent the horizontal and vertical luminance changes; and D represents the interocular distance. The derivation of equation (1) is provided in the Appendix. Intuitively, however, equation (1) shows that, for each image point, there is a one-to-one mapping between the luminance gradient and the disparity 'curvature' (ie the second-order derivative of the disparity field). It should be noted that equation (1) only holds for Lambertian shading, collimated parallel illumination, no inter-reflections, and no body or cast shadows. Under these conditions (which were simulated in the experiments reported), the shading gradient could constrain the range of possible disparity values in image regions where there is no disparity information.

The purpose of the present empirical investigation was to establish whether the theoretically possible cooperation between shading and stereo information to 3-D shape expressed by equation (1) might take place. In two experiments, we used a probe-adjustment task adapted from Lappin and Craft (2000) (see also Vuong et al 2004). In this task, observers adjust a probe dot defined by disparity to lie on a simulated convex surface defined either by disparity, shading, or both cues. Our goal in the present study was not to make quantitative predictions on the basis of equation (1). Rather, our purpose was to test the hypothesis that the visual system could use such constraints to determine disparity values in regions where no measurable disparities are present. If so, we expect observers to be less variable in the probe-adjustment task when a shading gradient is added to an otherwise sparse disparity field. In this study we focused on the variance of observers' judgments, rather than on their absolute depth estimation for comparison with previous studies on cue combination, and because Bülthoff and Mallot's (1988) results show that absolute perceived depth varied dramatically as a function of the type of depth cue and number of cues used.

2 Experiment 1

The purpose of experiment 1 was to directly test for cooperation between disparity and shading information for surface interpolation. Following previous work, we measured observers' precision (ie variance) at placing a probe dot on a surface specified only by stereo disparity, only shading, or a combination of both disparity and shading cues (eg Jacobs 1999).

The inclusion of a shading-only condition in experiment 1 was used as a control condition to address two issues. First, we could test whether the presence of a weak gradient-based disparity signal could cooperate with (monocular) shading cues for surface interpolation. Second, we wanted to discount the possibility that improvements in precision are simply due to the availability of more cues (Landy et al 1995). To do this, we used observers' variance in the single-cue conditions (ie disparity or shading) to predict their variance in the combined-cue conditions (eg Ernst and Banks 2002; Jacobs 1999):

$$\sigma_{\mathrm{DS}}^2 = \frac{\sigma_{\mathrm{D}}^2 \sigma_{\mathrm{S}}^2}{\sigma_{\mathrm{D}}^2 + \sigma_{\mathrm{S}}^2}\,, \tag{2}$$

where the subscripts denote the variance obtained in the disparity cue (D), the shading cue (S), or the combined cues (DS). If we assume independent estimates corrupted by Gaussian noise, equation (2) produces the lowest variance in the combined-cue estimate (Jacobs 1999; Landy et al 1995; Yuille and Bülthoff 1996). Any variance less than this predicted variance would reveal a cooperation between disparity and shading cues, and argue against the possibility raised above. The use of equation (2) allows us to compare and relate our results to previous studies on cue combination. However, this equation makes the assumption that the separate cues are used to derive a depth map (Landy et al 1995). This may not be a valid assumption, as shading information may be used to refine the estimation of shape, for example. We return to this important issue in section 4.

2.1 *Method*

2.1.1 *Participants.* The observers were two of the authors (QV and FD), and two female undergraduate students at Brown University (HK and MT). Both naive participants were unaware of the purpose of the study but were experienced psychophysical observers. All observers had normal or corrected-to-normal vision, and provided informed consent.

2.1.2 *Stimuli.* As illustrated in figure 1, the basic stimulus used in both experiments 1 and 2 consisted of a spherical surface with a rectangular base. The spherical part of the stimulus had a radius of 1.8°. For the purpose of the present study, only the spherical part of the stimulus was critical. The stimulus had a maximum simulated depth of 10.0 mm, corresponding to a maximum disparity of 185.5 s of arc. As described below, this surface could be defined by red dots, shading, or both.

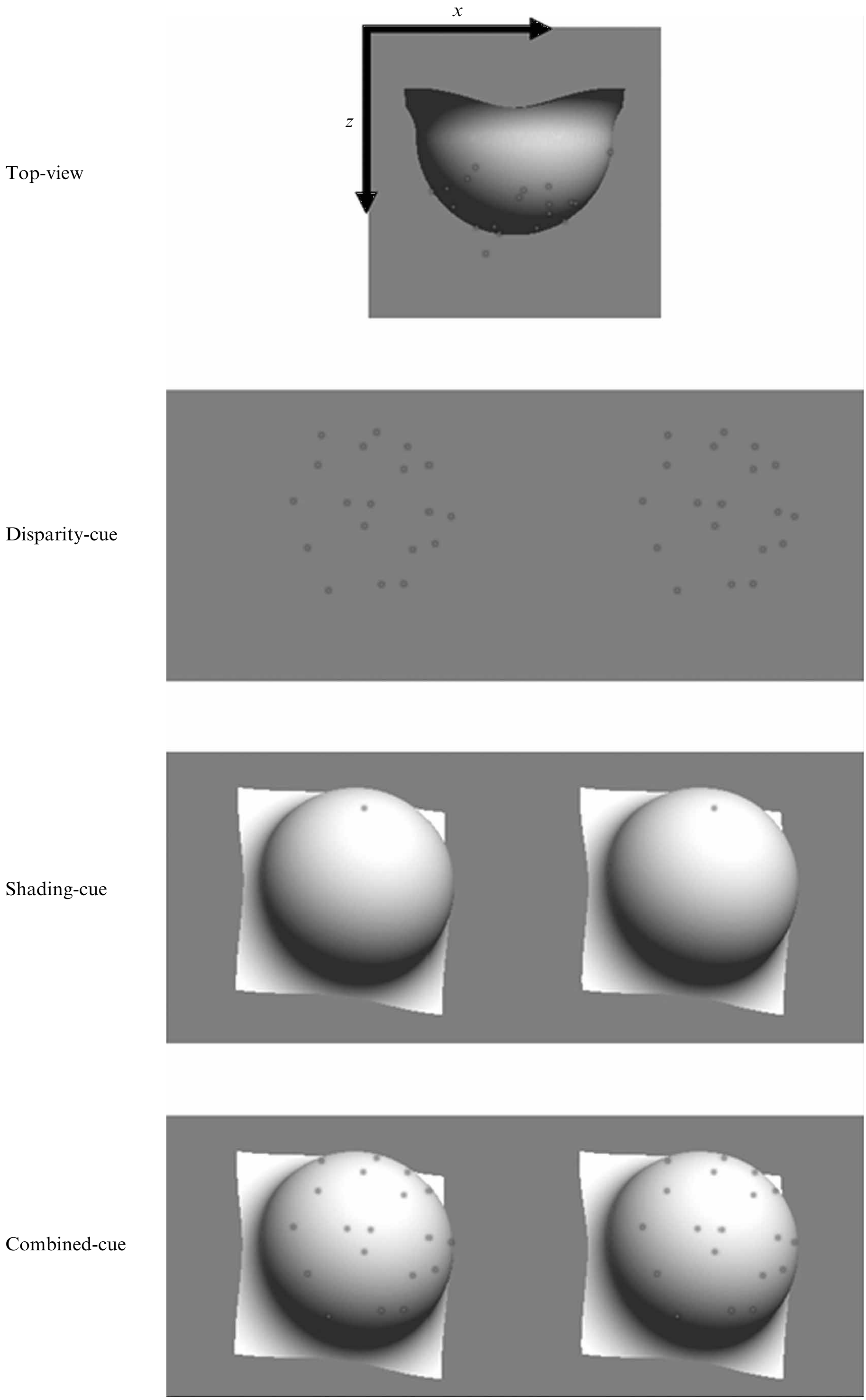

Figure 1. Stereograms of the stimulus display used in experiment 1, shown in the disparity-cue, shading-cue, and combined-cue conditions. Note that the dots are in the same configuration in these stereograms. When cross-fused, the probe dot should appear perturbed off the surface towards the viewer. A top view of the stimulus is also presented.

The surface was oriented such that when frontoparallel, observers saw a half-sphere centered on a rectangle. In experiment 1, the entire surface was rotated 20.0° up and to the right with respect to the observer (see figure 1).

2.1.3 *Design and procedure.* There were three presentation conditions, run within subjects. In the *disparity-cue* condition, the surface was defined by nineteen anti-aliased red dots ($\sim$100 cd mm^{-2}) on a dark-gray background (39 cd mm^{-2}). The dots were initially placed at 19 fixed locations that formed a hexagonal pattern on the image plane, and then they were randomly jittered vertically and horizontally by a maximum of 0.3 deg, with the constraint that the dots did not get perturbed beyond the radius of the sphere. The purpose of jittering the dots and randomly selecting a probe dot was to increase the variability of observers' adjustments from trial to trial. The distance between any neighboring pairs of dots on the image plane was, on average, 0.75 deg. The dots were then orthographically projected onto the spherical part of the otherwise invisible surface, and their disparity computed. The probe dot was randomly selected from one of six possible dots on a trial-by-trial basis.

In the *shading-cue* condition, we computed the luminance gradient for each monocular image assuming a Lambertian surface, orthographic projection, and a single light source. This stimulus provided both stereo shading and monocular shading cues. For purposes of selecting a probe location, the procedure described above was used to generate nineteen dots. As before, the probe dot was randomly selected from one of six possible dots on a trial-by-trial basis, but only this dot was rendered. Lastly, in the *combined-cue* condition, we superimposed the dots onto the luminance gradient of each image.

On each trial, the surface was presented and a probe dot randomly selected. The probe dot was jittered along the surface normal at the probed location by a random amount towards or away in depth from the observer. This jitter had a mean of 789.1 s of arc and a standard deviation (SD) of 92.1 s of arc. The observer adjusted the horizontal disparity of the probe dot until it was coincident on the surface, with the constraint that the probe dot was displaced along the normal. Four keys were designated for this adjustment: two of the keys moved the probe in small steps of $\pm$7.9 s of arc of disparity, and the remaining two moved the probe in large steps of $\pm$39.5 s of arc of disparity. Each key adjustment moved the probe along the surface normal at the probed location. When observers were satisfied that the probe dot was on the surface, they pressed a mouse button to advance to the next trial. There was no time pressure; each trial took approximately 2 – 3 min. Observers participated in four blocks of 30 trials. In each block, the disparity-cue, shading-cue, and combined-cue conditions were randomly presented. Across the four blocks there were 40 repetitions of each condition. Observers took long breaks between blocks.

2.1.4 *Apparatus.* In both experiments, the stimulus was presented binocularly. A calibrated mirror system was used to present left and right images to corresponding eyes. Observers sat approximately 1150 mm from a monitor with 1600 $\times$ 1280 addressable locations [see Vuong et al (2004), for details]. At this viewing distance the fused spherical part of the stimulus subtended approximately 3.6 deg $\times$ 3.6 deg of visual angle. The dots subtended approximately 0.03 deg $\times$ 0.03 deg. The experiments were conducted in a dark room.

2.2 *Results and discussion*

For experiments 1 and 2, our main analysis was based on the relative improvement in observers' precision in the disparity-cue and combined-cue conditions. On each trial, we measured observers' adjustment error, defined as the absolute difference between the estimated disparity of the probe dot and its simulated disparity when it is on the surface at the probed location. We used the standard deviation (SD) of these errors to quantify observers' precision in each presentation condition (Lappin and Craft 2000). Individual observers' performance in experiment 1 is shown in table 1. Our analyses focused on the relative improvement in precision given two estimates of precision, expressed as a percentage improvement (IMP) defined as:

$$\mathrm{IMP} = 100 \times \left(1 - \frac{\mathrm{SD}_1}{\mathrm{SD}_2}\right) . \tag{3}$$

For example, if observers are twice as precise in condition 1 (SD_1) relative to condition 2 (SD_2), then this would reflect a 50% improvement. One-tailed t-tests with $\alpha = 0.05$ were used to assess the statistical significance of the improvement averaged across observers (ie whether improvement is greater than or less than 0%).

Table 1. Standard deviations (SD) of adjustment errors in experiment 1. Also shown is the predicted SD from the MWF model; and the relative improvement between the observed combined-cue SD and predicted SD, and between the combined-cue SD and disparity-cue SD.

Subject	SD/s of arc				Improvement/%	
	disparity	shading	combined	predicted from MWF	combined to predicted	combined to disparity
FD	8.41	24.88	6.28	7.96	21.1	25.3
HK	11.01	23.98	7.90	10.01	21.1	28.3
MT	33.55	54.03	25.49	28.50	10.6	24.0
QV	5.51	38.99	3.82	5.46	30.0	30.7
M	14.62	35.47	10.87	12.98	20.7	27.1
(SEM)	(7.40)	(8.17)	(5.71)	(6.07)	(4.6)	(1.7)

In experiment 1 we found that observers were very precise at estimating the disparity of the probe dot in the disparity-cue condition ($M = 14.7$ s of arc, SEM $= 7.4$ s of arc), as previously found (Lappin and Craft 2000; Vuong et al 2004). By comparison, observers were much worse on this task in the shading-cue condition ($M = 35.5$ s of arc, SEM $= 8.2$ s of arc), which is also consistent with earlier results (Bülthoff and Mallot 1988). The critical finding in this experiment was the significant improvement in performance in the combined-cue condition ($M = 10.9$ s of arc, SEM $= 5.7$ s of arc): observers were 27.1% (SEM $=$ 1.7%) more precise in the combined-cue condition than in the disparity-cue condition ($t_3 = 18.1$, $p < 0.01$). We point out that the improvement found in the combined-cue condition relative to disparity-cue condition cannot be due to the fact that there are simply more cues available. We found that observers were 20.7% (SEM $=$ 4.6%) more precise in the combined-cue condition than the predicted SD [equation (2)] ($t_3 = 5.21$, $p < 0.01$), suggesting that there was an interaction between stereo and shading cues.

3 Experiment 2

The results from experiment 1 show that observers are more precise at interpolating the depth of a probed location on surface when both disparity and shading cues are present than when only disparity information is present. However, in that experiment, there were potential depth cues from stereo shading and from contours that could also contribute to observers' depth estimation. In experiment 2 we eliminated the weak disparity signal by presenting *identical* images to the left and right eye. We also presented the surface frontoparallel to eliminate any contour cues to depth (Tse 2002). Thus, in the shading-cue condition of experiment 2, the fused stimulus consisted of a disk with the luminance profile of a sphere 'floating' above a uniform-luminance rectangle.

There was a strong disparity signal at the contour of the disk because the disk and background had different depth; however, this disparity signal specified a flat disk as there was no disparity in the spherical region.

3.1 *Method*
3.1.1 *Participants.* The same two authors and a naive observer from experiment 1 (HK) participated in experiment 2.

3.1.2 *Stimuli.* The same stimulus as that used in experiment 1 was used in experiment 2. As described above, the exceptions were, first, that the same image was presented to the left and right eye, and, second, that the surface was presented frontoparallel. Figure 2 illustrates the stimulus used in experiment 2.

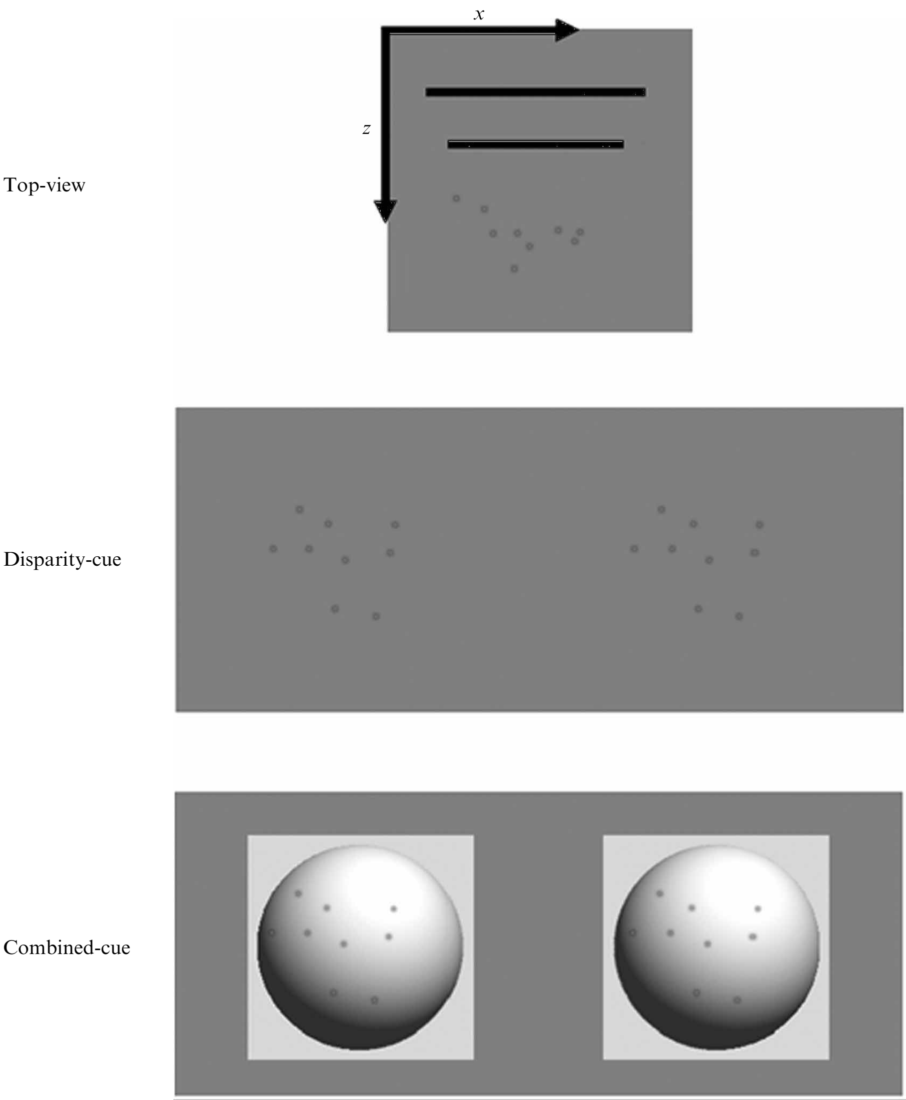

Figure 2. Stereograms of the stimulus display used in experiment 2, shown in the disparity-cue and combined-cue conditions. Note again that the dots are in the same configuration in these stereograms. When cross-fused, the probe dot should appear perturbed off the surface towards the viewer. In the top view of the stimulus display, the black bars indicate the location of the flat disk (short bar) and the rectangle (long bar) relative to the dots.

3.1.3 *Design and procedure.* All observers repeated the adjustment task with two modifications. First, only nine positions were randomly sampled on the implicit surface in the disparity-cue condition (with the same procedure as the one used in experiment 1). Again, one of these dots was randomly selected as the probe dot on each trial. Second, only the disparity-cue and combined-cue conditions were tested since pilot work indicated that observers found this task extremely difficult in the shading-cue condition. Observers participated in two blocks of 20 trials. In each block, the disparity-cue and combined-cue conditions were randomly interleaved. Across the two blocks, there were 20 repetitions of each condition. As in experiment 1, observers took long breaks between blocks.

3.2 *Results and discussion*

The results of experiment 2 replicate those of experiment 1, despite the fact that (i) only eight dots were used in the disparity-cue and combined-cue conditions, (ii) only monocular shading cues were available in the combined-cue condition, and (iii) the surface was presented frontoparallel (see figure 2). First, observers perform very precisely in the disparity-cue condition ($M = 14.9$ s of arc, SEM $= 2.7$ s of arc, see table 2). Second, performance also improved with the addition of monocular shading cues: we found an average of 30.6% (SEM = 8.6%) improvement in the combined-cue condition ($M = 10.6$ s of arc, SEM $= 3.3$ s of arc) relative to the disparity-cue condition ($t_2 = 4.34$, $p < 0.02$). Taken together, the results from experiments 1 and 2 suggest that shading cues, though not useful for the probe-adjustment task, could cooperate (ie interact) with disparity cues to constrain how disparities were estimated in regions where there were *no* disparity signals.

Table 2. Standard deviations (SD) of adjustment errors in experiment 2. Also shown is the relative improvement between the combined-cue SD and the disparity-cue SD.

Subject	SD/s of arc		Improvement/%
	disparity	combined	combined to disparity
FD	15.52	12.24	21.3
HK	18.39	13.54	26.3
QV	10.82	6.02	44.4
M	14.91	10.60	30.6
(SEM)	(2.70)	(3.32)	(8.6)

4 General discussion

The main aim of the present study was to test the extent to which disparity and shading cues cooperated for surface interpolation (see also Bülthoff and Mallot 1988). To address this problem, we measured how consistently observers placed a probe dot on a simulated surface which was specified by disparity, shading, or both cues. Overall observers were very consistent from trial to trial as reflected by the very small magnitude of deviations in their adjustments (in the s of arc range). In experiment 1, observers performed more precisely when both disparity and shading cues were available than when only a single cue was available. Moreover, there is evidence that observers could perform better than predicted by an optimal linear combination model [see equation (2)]. This finding suggests that a better performance in the combined-cue condition cannot be explained in terms of an optimal combination of independent 3-D estimates from disparity and shading cues.

In experiment 2, observers could not perform the task in the monocular-shading condition, but their performance significantly improved when monocular shading was added to the disparity cue. Previously we found a slight improvement in precision with a non-jittered hexagonal pattern relative to a jittered pattern, suggesting that regularity of the 2-D pattern might increase precision (Vuong et al 2004). The current findings demonstrate more systematically that observers also use shading information to reduce their variability.

Given the stimulus used in the present study (see figures 1 and 2), several additional factors may also contribute to the improvement found when both disparity and shading cues are available. First, in both the shading and combined conditions, there is a visible contour which, as raised above, can provide a strong depth cue (eg Tse 2002). At present, we do not know how to quantify the relationship between contour and disparity, as we have done for the relationship between luminance gradient and disparity [see equation (1)]. We note, however, that observers performed poorly in the shading-only condition in experiment 1 even though contours were available in that condition. This implies that any contributions from contour information may be small for this particular task, but future studies are needed to address this important issue.

Another potential depth cue that might have contributed to the results of experiment 1 was the weak disparity signal from stereo shading (see also Bülthoff and Mallot 1988). We partially addressed this issue in experiment 2 in which this weak stereo signal was eliminated by presenting identical images to each eye. That we still found an improvement in the combined condition relative to the disparity-only condition again suggests that any contribution from stereo shading may be small for the probe-adjustment task. In line with this argument, several computational and behavioral studies have shown that stereo shading is a weak cue to 3-D shape perception (eg Arndt et al 1995; Mallot et al 1996).

Overall, we believe that the results of experiments 1 and 2 are due to a cooperative processing of disparity and shading cues (Bülthoff and Mallot 1988). This proposal is rooted in the observation that isomorphic relations exist, in real-world situations, among different 2-D depth cues. In any 2-D projection of the natural environment, in fact, the image signals specified by different depth cues are, by necessity, linearly related. For example, there is a linear relationship between the luminance gradient and the second-order disparity gradient projected by a given point on a surface. Although the relationship expressed in equation (1) only holds for a particular set of conditions (Lambertian shading, no inter-reflections, no cast shadows, etc), we believe that similar isomorphic relations exist for other conditions and other depth cues. A goal of our current studies is to examine other isomorphic relationships between 2-D cues to depth. In any case, such constraint concerning the properties of the retinal projections has only been hinted at, but never studied as such. Qian and his colleagues, for example, derived a biologically plausible model of stereo-motion integration from the observation that many cells in visual areas have both disparity and velocity tuning (eg Fernandez et al 2002; Qian 1994; Qian and Andersen 1997).

This proposal departs from several strong assumptions in the MWF model (Landy et al 1995). As raised in section 1, the critical assumption is that visual information is used to create a depth map in which a depth value is assigned to each point in an image. Given this assumption, it is further assumed that different depth cues lead to different depth estimates, and that these independent cues are then linearly combined. In effect, surfaces are derived from the relative depths at image points. By comparison, we believe that the primitive visual information includes surfaces and their shapes rather than depth estimates (eg Lappin and Craft 2000; Vuong et al 2004). Therefore, in contrast to the MWF model, relative depth values are derived from surfaces. Under natural viewing circumstances, different depth cues are necessarily highly correlated because they provide information about the same surface. The results of the present study strongly suggest that observers could use this constraint to improve their performance.

In conclusion, we favor the possibility that the highly precise performance we found may be attributed to a process that capitalizes on constraints derived from the natural covariation among 2-D depth signals. Here we have demonstrated that human observers take advantage of these constraints.

Acknowledgments. This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) post-graduate scholarship to QV and a National Science Foundation grant 78441 to FD. We would also like to thank two anonymous reviewers for their helpful suggestions and comments.

References

Arndt P A, Mallot H A, Bülthoff H H, 1995 "Human stereovision without localized image features" *Biological Cybernetics* **72** 279 – 293
Bülthoff H H, Mallot H A, 1988 "Integration of depth modules: Stereo and shading" *Journal of the Optical Society of America A* **5** 1749 – 1758
Clark J J, Yuille A L, 1990 *Data Fusion for Sensory Information Processing* (Boston, MA: Kluwer Academic)
Ernst M O, Banks M S, 2002 "Humans integrate visual and haptic information in a statistically optimal fashion" *Nature Neuroscience* **415** 429 – 433
Fernandez J M, Watson B, Qian N, 2002 "Computing relief structure from motion with a distributed velocity and disparity representation" *Vision Research* **42** 883 – 898
Horn B K P, 1986 *Robot Vision* (Cambridge, MA: MIT Press)
Jacobs R, 1999 "Optimal integration of texture and motion cues to depth" *Vision Research* **39** 3621 – 3629
Johnston E B, Cumming B G, Landy M S, 1994 "Integration of stereopsis and motion shape cues" *Vision Research* **34** 2259 – 2275
Landy M S, Maloney L T, Johnston E B, Young M, 1995 "Measurement and modeling of depth cue combination: In defense of weak fusion" *Vision Research* **35** 389 – 412
Lappin J S, Craft W D, 2000 "Foundations of spatial vision: From retinal images to perceived shapes" *Psychological Review* **107** 6 – 38
Mallot H A, Arndt P A, Bülthoff H H, 1996 "A psychophysical and computational analysis of intensity-based stereo" *Biological Cybernetics* **75** 187 – 198
Qian N, 1994 "Computing stereo disparity and motion with known binocular cell properties" *Neural Computation* **6** 390 – 404
Qian N, Andersen R A, 1997 "A physiological model for stereo-motion integration and a unified explanation of Pulfrich-like phenomena" *Vision Research* **37** 1683 – 1698
Tse P U, 2002 "A contour propagation account of surface filling-in and volume formation" *Psychological Review* **109** 91 – 115
Vuong Q C, Domini F, Caudek C, 2004 "Evidence for patchwork approximation of shape primitives" *Perception & Psychophysics* **66** 1246 – 1259
Yuille A L, Bülthoff H H, 1996 "Bayesian decision theory and psychophysics", in *Perception as Bayesian Inference* Eds D C Knill, W Richards (Cambridge: Cambridge University Press) pp 123 – 161

Appendix

In this Appendix we derive equation (1), which shows the linear relationship between the luminance gradient and the second-order disparity gradient. Let us consider a coordinate system (x, y, z) where z is the depth dimension and (x, y) is the image plane. If a Lambertian surface is illuminated by a distant light source then it can be shown that:

$$\begin{bmatrix} \dfrac{\partial g_x}{\partial \xi} \\ \dfrac{\partial g_y}{\partial \xi} \end{bmatrix} = \begin{bmatrix} I_x \\ I_y \end{bmatrix}, \tag{A1}$$

where g_x and g_y are the horizontal and vertical depth gradients; I_x and I_y are the horizontal and vertical luminance gradients; and ξ is the direction of maximum luminance change on the image (Horn 1986). If an observer is viewing a local surface $z(x, y)$ at a distance z_f from the image plane (x, y) then the disparities (described in terms of visual angles) projected by the visible points on the surface can be approximated by the following equation:

$$d = \frac{D}{z_f^2} z(x, y), \tag{A2}$$

where D is the interocular distance. We indicate with (α, β) the angles subtended by a point $P(x, y)$ on the image plane. If the surface is local (ie it subtends a small visual angle) then

$$\alpha \approx \frac{x}{z_f} \text{ and } \beta \approx \frac{y}{z_f}.$$

If we now derive the disparity field with respect to α and β we obtain:

$$d_\alpha = \frac{D}{z_f^2}\frac{\partial z}{\partial \alpha} = \frac{D}{z_f^2}\frac{\partial z}{\partial x}\frac{\partial x}{\partial \alpha} = \frac{D}{z_f^2} g_x z_f = \frac{D}{z_f} g_x$$

and similarly:

$$d_\beta = \frac{D}{z_f} g_y \,.$$

If we now indicate with γ the angular subtend in the direction ξ $\left(\text{so that } \gamma \approx \dfrac{\xi}{z_f}\right)$ then the derivative of the disparity gradient with respect to γ is:

$$\frac{\partial d_\alpha}{\partial \gamma} = \frac{\partial d_\alpha}{\partial \xi}\frac{\partial \xi}{\partial \gamma} = \frac{D}{z_f}\frac{\partial g_x}{\partial \xi} z_f = D\frac{\partial g_x}{\partial \xi}$$

and similarly:

$$\frac{\partial d_\beta}{\partial \gamma} = D\frac{\partial g_y}{\partial \xi}.$$

Since from equation (A1) $\dfrac{\partial g_x}{\partial \xi} = I_x$ and $\dfrac{\partial g_y}{\partial \xi} = I_y$ we finally obtain:

$$\begin{bmatrix} \dfrac{\partial d_\alpha}{\partial \gamma} \\ \dfrac{\partial d_\beta}{\partial \gamma} \end{bmatrix} = D\begin{bmatrix} I_x \\ I_y \end{bmatrix}, \tag{A3}$$

that shows a linear relationship between the second-order disparity field and the luminance gradient.

ISSN 0301-0066 (print) ISSN 1468-4233 (electronic)

PERCEPTION
VOLUME 35 2006

www.perceptionweb.com