

The combined influence of binocular disparity and shading on pictorial shape

PAUL C. A. DOORSCHOT, ASTRID M. L. KAPPERS, and JAN J. KOENDERINK
Universiteit Utrecht, Utrecht, The Netherlands

The combined influence of binocular disparity and shading on pictorial shape was studied. Stimuli were several pairs of stereo photographs of real objects. The stereo base was 0, 7, or 14 cm, and the location of the light source was varied over three positions (one from about the viewpoint of the camera, one about perpendicular to the line of sight, and one in between the two). Therefore, in total, nine different combinations were studied. Subjects had to perform surface attitude settings at about 300 positions in the image plane. From the settings, depth maps were calculated on which a principal components analysis was performed. It was found that three components were enough to account for at least 97.8% of the variance in the data. The first component accounted for shape constancy. The effects of the two cues could be isolated as a linear combination of the other two components. The effects of the disparity and the shading cue variation were found to combine in almost linear fashion.

Ecological optics reveals various aspects of a scene or a picture of a scene as potential *depth cues* or *shape cues*, such as texture gradients, binocular disparity, contour, and shading. A large number of studies have been carried out on the role of single cues in human perception. In real life, typically many cues are available to an observer simultaneously, so it is important to proceed from the study of single cues to the case of combined cues. This is a complicated issue because of the varied nature of the cues. Several studies have focused on this subject.

In our experiment, we addressed the combination of binocular disparity and shading as the cause of pictorial relief. These are two cues whose individual effects have been well researched. In the following paragraphs, we argue that these cues provide data that are potentially complementary in many respects, which makes it interesting to see how they combine. Another important reason why these particular cues have been chosen is that they can be varied parametrically.

First, we consider the shading cue. We use the term *shading* to refer to a number of aspects of surface scattering, all of which may provide potential shape information. For many surfaces, surface scattering can be decomposed roughly into a specular component (regular reflection) and a Lambertian component. The specular component causes highlights; the Lambertian component leads to radiance variations when the local surface attitude varies with respect to the major direction of irradiation. In addition, the light source or part of it can be shielded from

the surface by another surface part. In the case of a collimated source, this yields attached and possibly cast shadows, and when the light source is extended, one speaks of vignetting. The transition between cast shadows and vignetting is a gradual one.

We would like to point out that the shading cue is not of a global nature, because (generally) the irradiating beam is not uniform over the whole scene. For instance, even if we look at a simple object that is illuminated by just one source, the light comes from a different direction for parts that are illuminated directly, as compared with parts that are in the shadow. Furthermore, a linear radiance gradient is compatible with any type of constant curvature, so it can for instance be both a concave and a convex object that caused an exactly identical gradient. Ramachandran (1988) argued that the shading cue is used on the assumption that the whole scene an observer sees is irradiated from only one direction. Even if that is indeed the case, shading patterns can still vary tremendously, depending on the exact location of the light source. At the same time, a given specific shading pattern can be generated by many different shapes. So, unless an observer has access somehow to source direction and albedo variation, the shading cue is of an ambiguous nature.

Second, we consider the binocular disparity cue. For the visual system to be able to use this as a cue, surfaces need to be revealed by well-localized surface markings (see, for instance, Blake, Zisserman, & Knowles, 1985; Marr, 1982). Potential features are edges, texture, and shading. However, straightforward fusion of smooth edges on smooth objects, like, for instance, a sphere, leads to incorrect disparity, because, due to self-occlusion, the edge in the right image originates from a different part of the sphere than does the edge in the left image (Todd, Norman, Koenderink, & Kappers, 1997). Fusion of dihedral edges, like those of a cube, however, leads to correct dis-

This research was supported by the Life Sciences Foundation (SLW), which is subsidized by the Netherlands Organization for Scientific Research (NWO). Correspondence concerning this article should be addressed to A. M. L. Kappers, Helmholtz Instituut, Universiteit Utrecht, Princetonplein 5, 3584 CC Utrecht, The Netherlands (e-mail: a.m.l.kappers@phys.uu.nl).

parity. By *texture* we mean variation of the surface albedo, rather than three-dimensional (3-D) surface roughness. Like dihedral edges, texture is well localized and therefore leads to correct disparity. Radiance variation due to attached and cast shadow boundaries also creates possibilities for useful binocular matches. The specular component leads to matches whose disparities do not reveal the surface in a simple way, because the spatial structure of this component depends on the vantage point.

In the present experiment, we varied the disparity cue parametrically by varying the stereo base from zero to about twice the natural stereo base in pairs of stereo photographs. When the stereo base is varied, three parameters will be altered: the absolute disparity and with it the relative disparity (see Blakemore, 1969, for the distinction between the two) and the vergence. It is important to note that a change of any of these parameters will *globally* affect the percept of a scene. Collewijn and Erkelens (1990) argued that any role played by vergence-related signals or absolute disparity in the estimation of distance can only be weak: Under full cue conditions, other cues dominate. However, binocular stereopsis is sensitive to *relative* disparities, and relative disparity is an effective cue to relative depth differences. From geometrical considerations, it can be concluded that, when the stereo base is altered, relative disparity will signal a differently scaled disparity field. Therefore, the same object viewed with a higher stereo base might be interpreted as being more elongated in depth (see also Howard & Rogers, 1995, for a theoretical discussion on stereopsis). On those grounds, we expect *global* scaling effects to occur when the stereo base is varied. This expectation is based only on the assumption that the visual system presumes a constant stereo base, which indeed seems to be the case (Howard & Rogers, 1995).

In the past, such scaling effects have indeed been found in binocular viewing experiments in which the stereo base was varied (Koenderink, van Doorn, & Kappers, 1995). The natural stereo base of the subjects and stereo base zero were used. It was found that, in the zero disparity condition, shapes were globally judged to be flattened. However, in these experiments, unlike in the present experiment, subjects looked at a real scene instead of photographs. Note that, in the stereo base zero condition, relative disparity signals a disparity field with zero depth, so, if the observer had based shape judgments on disparity alone, shapes would have been judged to be totally flat, which was not the case.

We mentioned that the disparity and the shading cue provide complementary shape information and depend on different prior assumptions. We pointed out that the shading cue is not of a global nature, and it also has an ambiguous nature. In contrast, the disparity cue is of a global nature. When a scene is illuminated by, for instance, two light sources, local parts of the scene may be illuminated by only one source. However, there is no way that we can, for instance, have a real scene with one stereo base on one part and another one on another part. Another difference is the fact that shading variations depend

mainly on the surface curvature, but disparity depends only on the distance. For instance, if we look at a uniformly illuminated flat slanted surface in isolation, shading provides us with no information whatsoever about this slant (or the depth). In that case, disparity is still available.

We mentioned that shading can help to provide disparity information by means of, for instance, attached shadow boundaries. On the other hand, disparity information can be used to disambiguate shading (e.g., Blake et al., 1985). So, the two cues can be complementary. Therefore, it will be very interesting to see whether and, if so, how these two cues are combined and whether effects of disparity variation depend on the lighting of the scene.

Bülthoff and Mallot (1988) addressed the issue of cue interaction between disparity and shading on simple objects (ellipsoids) with computer-generated shading. They found that the amount of elongation in depth of judged surfaces was an accumulation of the contribution of the two cues. They also found that, despite the smooth nature of radiance variation due to the Lambertian component, this could still be used as a feature for useful binocular stereopsis. A major difference between our experiment and the experiment of Bülthoff and Mallot is that we use a much more complex stimulus.

Landy, Maloney, Johnston, and Young (1995) reviewed the literature on cue combination and interaction and presented a formal model of depth cue combination that is consistent with many experimental results. The main purpose of this model is not to describe how the visual system works, but rather to guide cue combination experiments. The model assumes a depth system that is divided into different modules for different depth cues. In each module, a depth map is "computed," and also the reliability of this depth map is estimated. This reliability can vary between positions in a scene. The various depth estimates are combined into a single depth map by a weighted linear average, where the weights take into account the estimated reliability and also possible discrepancies between depth maps. By also accounting for these discrepancies in the final depth estimate, Landy et al. (1995) stressed that their model becomes *robust*, which means that, if the deviation of one cue, as compared with the other cues, increases from zero, it should affect the final depth estimate linearly. However, when the deviations become larger and increase beyond the range present in normal scenes, the influence of this one discrepant cue is assumed to have less and less effect on the percept. In the most extreme case, a cue can even be vetoed in this model if the information from this cue is too different from the information supplied by other cues.

Next, we bring up some issues of methodology. A common way to investigate the efficacy of a single cue is to try and isolate it from all the others. In most cases, however, one obtains ambiguous results. For instance, in the case of shading, subjects fail to obtain a clear impression of depth when other cues are totally absent (e.g., Erens, Kappers, & Koenderink, 1993). An alternative is to vary a single cue parametrically, while all others are kept constant. Such a *perturbation analysis* has been used success-

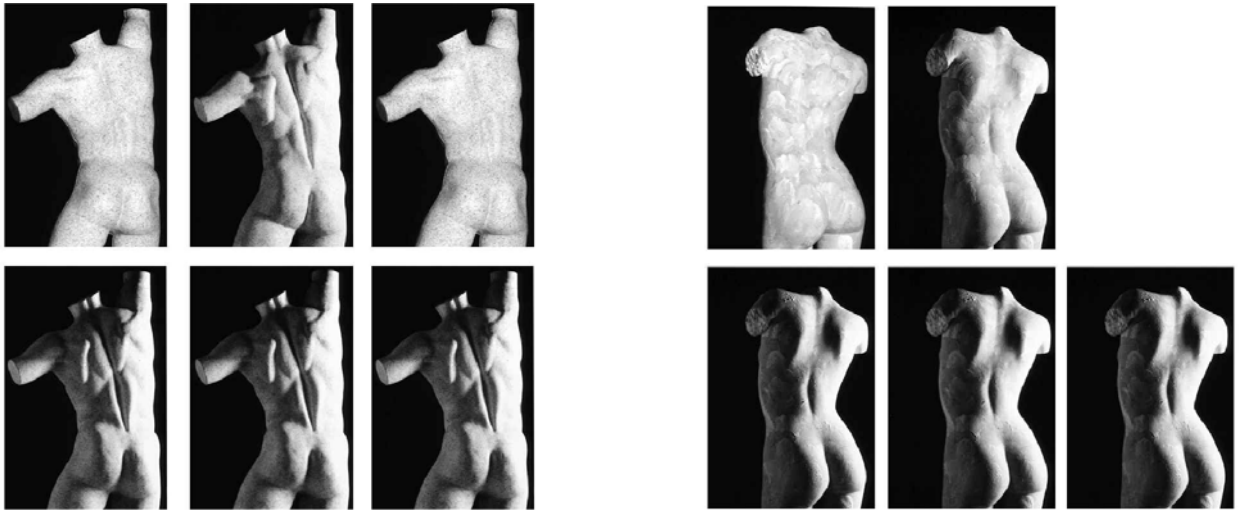


Figure 1. The stimuli used in the experiment. In A, B, and C, the photographs for stimulus A are shown for the three different lighting conditions. The photographs in A, B, and C were taken with the light source positioned about 1 m above the camera, at angles of 17.6°, 78.5°, and 101° to the line of sight, respectively. In C and D, the (uncrossed) stereo pair of photographs of this last lighting condition (rim) is shown, with a stereo base of 14 cm. A copy of C is also displayed on the right side of D, for crossed fusers. In E, F, and G, the photographs are shown for the same lighting conditions, but now they depict stimulus B. The stereo pair displayed in G and H has a stereo base of 7 cm. Again, a copy of G is also displayed on the right side of H, for crossed fusers.

fully in the case of shading by, for instance, Koenderink, van Doorn, Christou, and Lappin (1996a, 1996b) and Todd, Koenderink, van Doorn, and Kappers (1996). This methodology will therefore be applied in the present experiment. Perturbation studies, but of a different kind, have also been used to measure cue combinations or interactions (see, e.g., Johnston, Cumming, & Parker, 1993; Landy, Maloney, & Young, 1991; Young, Landy, & Maloney, 1993). In the latter experiments, conflicting cues were presented, and the weights based on the model of Landy et al. (1995) were measured. A disadvantage of cue conflicts is that the visual system is forced to operate in artificial and ambiguous situations, and the results do not necessarily apply to natural conditions.

Various paradigms address the perception of spatial structures. For instance, one may ask for the sign of elliptic curvature (i.e., convex or concave; Ramachandran, 1988), for the magnitude of the curvature (Buckley & Frisby, 1993), whether a line is perpendicular to a surface (Stevens & Brookes, 1988), and so forth. In such tasks, only simple, local shape aspects are to be judged, although the result may well depend on the global image. We used a very simple task for local slant and tilt (Koenderink, van Doorn, & Kappers, 1992). This task has been used extensively (e.g., see Koenderink et al., 1996a, 1996b; Koenderink et al., 1992, 1995; Norman, Todd, & Phillips, 1995; Todd et al., 1996; Todd et al., 1997). The advantages are as follows: A vast amount of data of geometrical nature can be gathered in a short amount of time; when desired, a global depth map of the investigated shape can be derived; almost no training is required to perform the task, and, even over longer periods of time, subjects are constant in their settings. Furthermore, subjects report that

the task comes to them naturally and they are certain of their settings. This task lends itself very well for perturbation studies in more natural, complicated scenes.

We present results of an experiment in which we addressed the issue of whether and, if so, how the disparity cue and the shading cue are combined in a natural setting. We varied these cues parametrically. In order to present realistic shading, we avoided computer graphics and used photographs. Stimuli were plastic human torsos that had very intricate shapes (see Figure 1). The model developed by Landy et al. (1995) predicts that the cues will combine linearly. If that prediction turns out to be true, we would expect to reproduce effects that are known or expected from studies of the cues in isolation: local effects for the shading cue, and more global effects for the disparity cue.

METHOD

Stimuli

Two plastic torsos were used as stimuli: one representing a male body (stimulus A), the other a female body (stimulus B). The surfaces were textured. Both objects were photographed in dorsal view (see Figure 1). Pairs of stereo photographs were presented to the subjects. The stereo bases were 0, 7, and 14 cm. The left photograph in these sets was always the same photo for the two stimuli.

The objects were photographed under three different illumination conditions (see Figure 2 for a set-up of the photo studio). In all cases, the camera was positioned 2.45 m from the stimulus. We used a light source with a directed diffuse beam (halogen light bulb in umbrella reflector of about 90-cm diameter), which was positioned about 1 m higher than the camera, at angles to the line of sight of 17.6° (the front lighting condition), 78.5° (the side lighting condition), and 101° (the rim lighting condition) and at distances of 2.88, 1.46, and 1.41 m from the stimulus, respectively. In Figure 1A, B, C, E, F, and G, the different illumination conditions are shown for the two stimuli. Figures 1C and 1D show the uncrossed fusion

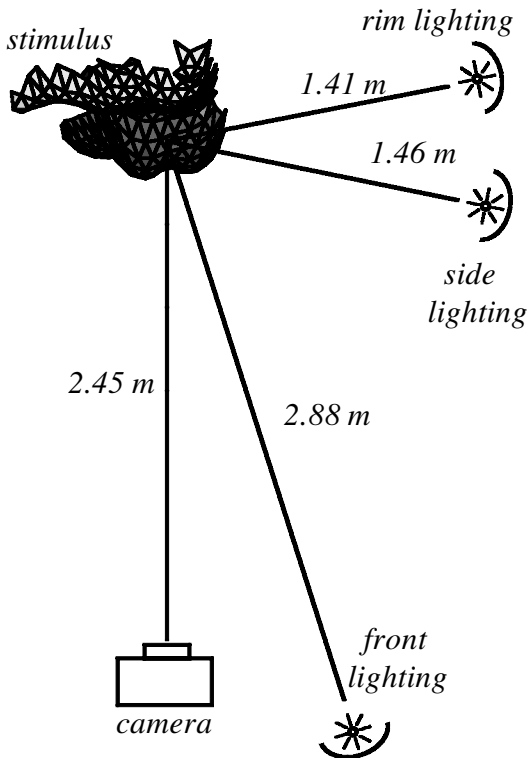


Figure 2. Schematic top view of the setup of the photo studio.

stereogram with a stereo base 14 cm for stimulus A in the rim lighting condition, and Figures 1G and 1H show a stereogram for stimulus B in the rim lighting, with stereo base 7 cm. Copies of Figures 1C and 1G are also displayed on the right side of Figures 1D and 1H, respectively, for people who prefer crossed fusion stereograms.

So, in total there were two objects, 3 disparity conditions, and 3 lighting conditions, which makes 18 different conditions in all.

Presentation of the Stimuli

The experiment was performed on a Quadra 950 Macintosh Computer with a PowerPC card. Two monitors were used, one color monitor (30.5 × 40.5 cm), on which the stimuli were presented, and one gray-scale monitor (17.5 × 22.5 cm), on which an interaction box was shown. This box was used for user interactions and cuing messages. The pictures were scanned with a Hewlett-Packard Scanjet plus scanner, which produced very high quality pictures on the monitor: comparable to good postcard-sized photographs (resolution: 28.3 pixels per cm; stimulus A: 347 × 520 pixels; stimulus B: 359 × 520 pixels). The pictures were viewed through a standard mirror stereoscope with a convergence angle of zero. In reality, the objects had a height of 89 cm, and the photographs were taken from a distance of 2.45 m. The height of the objects on the monitor was 18.5 cm; so, to look at them from the correct perspective, the viewing distance should be 55 cm. This was accomplished by fixing the stereoscope at the right distance. In that case, there were no cue conflicts between the vergence and the disparity cue. The room in which the experiments were performed was dimly lit, so that the outlines of the monitors were dimly visible.

Procedure

On the object in the left photograph, a red (monocular) wire-frame gauge figure was superimposed, which the subject could manipulate (see Koenderink et al., 1992). This gauge figure looked

like the orthographic projection of a circle (diameter ≈ 8 mm), with a stick perpendicular to the plane of the circle protruding from the center (length ≈ 4 mm). The task for the subject was to manipulate this gauge figure until it looked like a circle painted on the surface of the object, the stick being the outward surface normal. In this way, the gauge helped to resolve the inherent 180° ambiguity in the tilt. The gauge figure was presented monocularly to prevent the subjects from matching local disparities of the object and the gauge figure. The subjects saw the probe attached to the surface. The gauge figures were subsequently presented in random order on the vertices of a triangulation. For this purpose, the object in the left photograph was triangulated with a regular grid. Since all left pictures were the same, only two different triangulations were used: one for stimulus A, and one for stimulus B. Stimulus A was triangulated into 272 vertices, stimulus B into 265. The subjects never saw the triangulation during the experiments. An example of the triangulation for stimulus A is given in Figure 3.

In one session, which typically took about 1 h, the subjects were presented with 1 of the 18 conditions and had to adjust the gauge three times on all the vertices, in a randomized order. For 4 selected conditions, in eight extra sessions, the subjects performed an additional six trials per vertex. These 4 conditions were: stimulus A, with stereo base 0 cm, photographed with both the front and the rim lighting condition, and stimulus A, with stereo base 14 cm, also with both front and rim condition. Measuring every condition nine times would have been too time consuming, but we did want to get a better impression of the statistics than one would get with only three measurements per condition. That is why we selected 4 conditions to be measured nine times. These 4 conditions were chosen

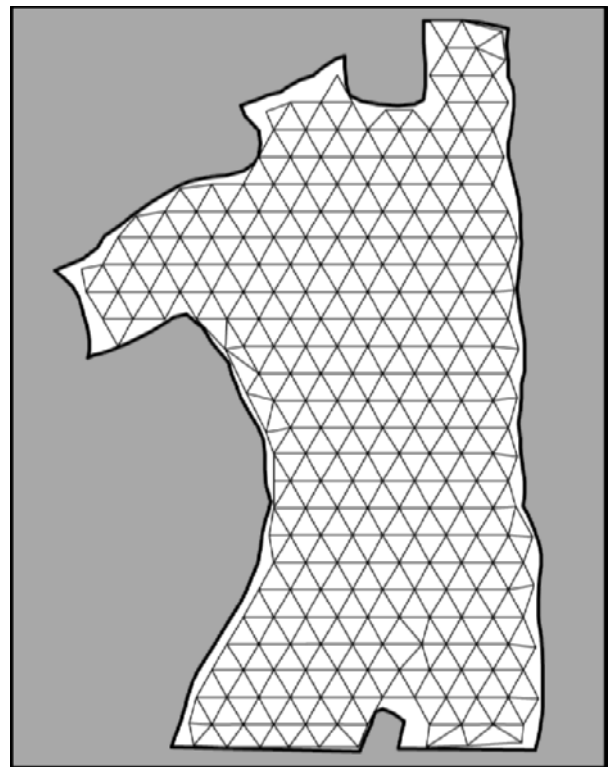


Figure 3. Example of the triangulation of stimulus A. Gauge figures were only presented on the vertices. Subjects never saw this triangulation during the experiment. The triangulation consisted of 272 vertices.

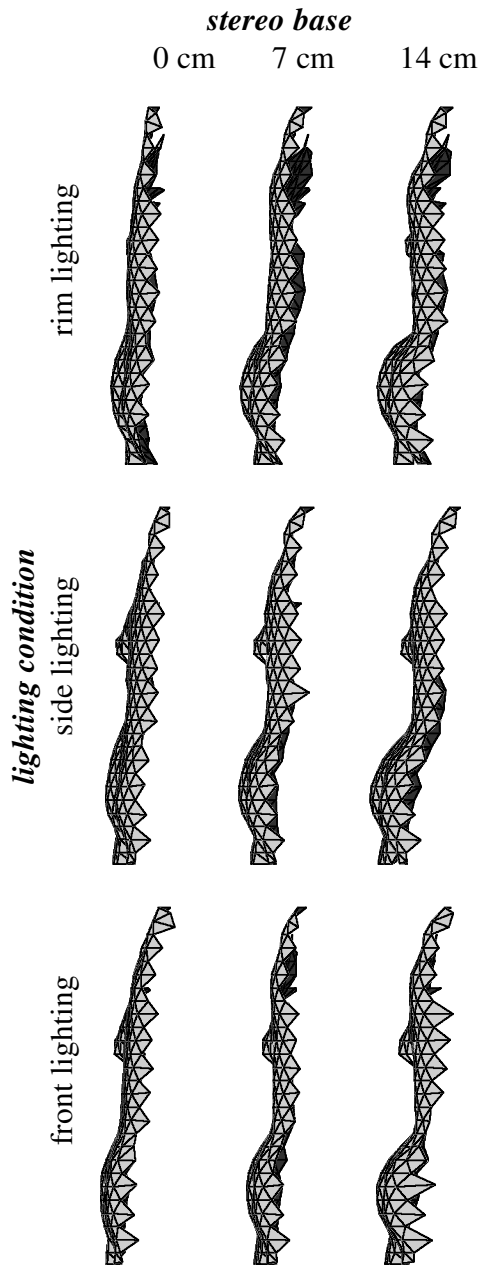


Figure 4. Side views of the constructions from the judgments of Subject B.Z. The leftmost column was measured with stereo base 0 cm, the middle with 7 cm, and the right with 14 cm. The upper, middle, and bottom rows were measured with lighting condition rim, side, and front, respectively.

because they were the most extreme parameter conditions. So, in total, there were 26 sessions, which were presented in randomized order. Since subjects typically performed one session per day, the whole experiment lasted for several weeks.

Subjects

Three naive, paid subjects (B.Z., R.H., and R.S.) and 1 nonnaive subject (P.D., the first author) performed the experiment. All had normal or corrected-to-normal acuity and good binocular vision, as verified with a TNO test (TNO, 1972).

Results

Constructions. We constructed surfaces from the subjects' settings to get an intuitive feeling for the data. Basically, the process came down to fitting the triangulation of a smooth surface to the local attitude settings. Thus we obtained a depth values map of about 300 points of the triangulation. Details about such a process can be found elsewhere (see Koenderink et al., 1992). These depth values maps can be depicted as 3-D surfaces (shapes or reliefs). Representative examples are shown in Figure 4 as side views for the nine surfaces of stimulus A for Subject B.Z. These are ordered in the horizontal direction by increasing the stereo base, in the vertical direction, by varying the lighting condition. Clearly these surfaces are similar, and they resemble a side view of the object that was photographed.

We noticed some trends in these side views: Constructions with stereo base 14 cm appeared more elongated in depth, whereas, with stereo base 0 cm, they appeared flatter. This can be seen in Figure 4. Another (less obvious) trend was that, with a higher stereo base, the angle between the lower back and the buttocks became more pronounced. Also, top views were constructed. In these, we noticed that in a few cases the upper part of the body looked twisted as compared with the bottom. Figure 5 shows an example of this for Subject P.D., with stereo base 14 cm and side and rim lighting. It can easily be seen in this figure that the angles between the "shoulder lines" (lines a and c) and the "buttocks lines" (lines b and d) are different in the two cases shown. Another trend we noticed in the top views for stimulus A was that the left shoulder/upper arm looked more curved in the side and rim lighting conditions. Though not as obvious as the former trend, this can also be seen in Figure 5. A local trend that we noticed in the top views for stimulus B was that the transition between the left and right buttock became more conspicuous in the side and rim lighting conditions.

Analysis of Results

Scatter plots. To further investigate relations between the different conditions, scatter plots of the depth values maps were constructed. In order to get a statistically well-balanced set of depth values, we only used the first three settings of the conditions that were measured nine times for the construction of scatter plots. Later, we also used these first three settings for the principal components analysis. In the scatter plots, depth values were plotted for two conditions, either with a different *disparity condition* but everything else (lighting condition, stimulus, and subject) constant, or with a different *lighting condition* and everything else constant. In the remainder of this paper, we will refer to these scatter plots as the *disparity* and the *lighting* scatter plots, respectively. In the disparity scatter plots, those values measured with the largest of the two stereo bases were always plotted along the y-axis. Similarly, in the lighting scatter plots, we plotted the depth values obtained in the condition with the largest angle to the line of sight along the y-axis. A lin-

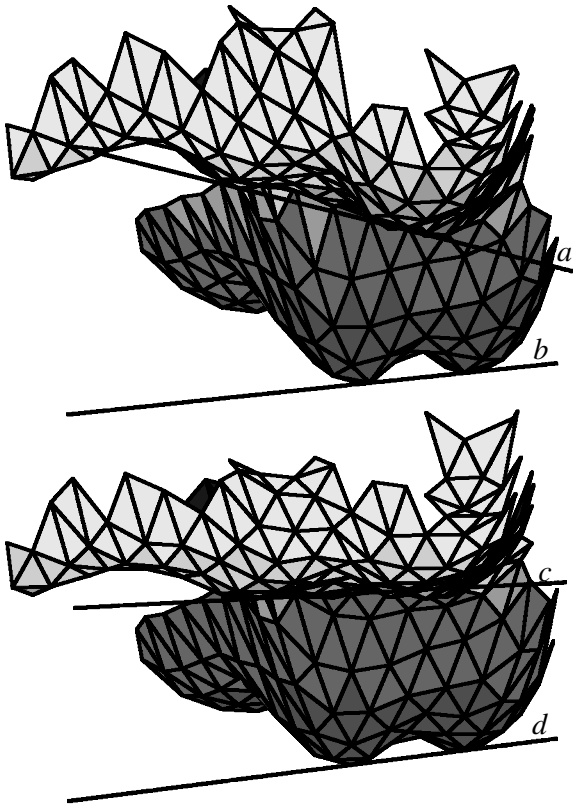


Figure 5. Two top views of the constructed surfaces for Subject P.D. The upper and bottom figures were photographed with lighting condition side and rim, respectively. Note how, in these top views, the shoulder blades look twisted in one plot with respect to the other.

ear regression was performed on the data. In Figure 6, an example of the disparity scatter plot is displayed for Subject B.Z., with rim lighting condition and stereo base 14 cm against stereo base 0 cm.

With a linear regression, the sum of squares of the error along the y-axis is minimized. Since in our plots no individual axis was preferred over the other, for the slopes, we minimized the sum of squares along the line *perpendicular* to the fitted line. As a measure for the goodness of the fit, we used the percentage of variance accounted for by the linear regression. These percentages were high: Over 90% were in the 93%–99% range, with a median of about 97%. In Table 1, mean slopes and mean goodness-of-fit values averaged over disparity and lighting scatter plots are presented for both stimuli. Slopes that differ significantly from 1.0 (two-sided students *t* test, $p < .05$, $df = 8$, $t_s > 2.306$) are marked with an asterisk. Some trends can be seen: Mean slopes in the disparity scatter plots were systematically higher than 1.0; in the lighting scatter plots, this was not the case. Furthermore, there were differences between subjects: especially Subject R.S., who deviated from the others in the slopes and also in the goodness-of-fit values for disparity, which were lower for R.S. than for the other subjects. In the remainder of our

analyses, it turned out that results for Subject R.S. typically differed from the results for the other subjects.

A number of different tests were carried out on the slopes and goodness-of-fit values. However, these tests did not lead to strong conclusions. Therefore, we present only a rough summary of the results of some tests. Only for stimulus A, which was measured nine times for four conditions, were data sufficient for further statistical claims. For Subject R.S., variation in the goodness-of-fit values could be accounted for by random scatter, but for the other 3 subjects this was not the case, which indicates that, for them, the cues did indeed have significant effects. However, the other tests we performed on the scatter plots did not clearly reveal what these effects were. Therefore, in the following section we describe another analysis technique that we used to further investigate the effects of the two cues. After this analysis, it will become evident why the scatter plots did not reveal these effects clearly.

Principal components analysis. The initial analysis did not contradict the idea that the relief depended in a continuous fashion on the parameter values. When only a single parameter was varied, the relief appeared to vary monotonically with the parameter value in complicated ways (e.g., local effects), differing for the lighting cue

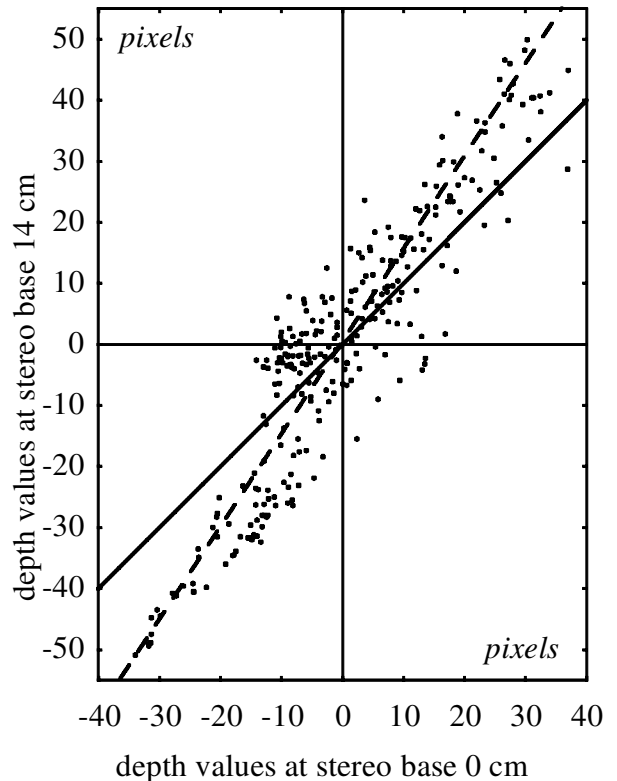


Figure 6. A scatter plot for Subject B.Z. On the x-axis we displayed the condition with stereo base 0 cm and lighting condition rim, on the y-axis we displayed stereo base 14 cm and also lighting condition rim. Also plotted are the lines $y = x$ and the best fitting line (slope = 1.51, goodness of fit = 96%). The best fitting line is displayed as a dotted line.

Table 1
Mean Slopes and Mean Goodness-of-Fit Values
of the Fits in the Scatter Plots for Each Subject
for Stimuli A and B, Split for the Two Cues

Subject	Mean R^2 Values and Mean Slopes							
	Mean Slope				Mean Goodness-of-Fit Values (%)			
	Disparity		Shading		Disparity		Shading	
	Stim A	Stim B	Stim A	Stim B	Stim A	Stim B	Stim A	Stim B
B.Z.	1.16*	0.96	1.02	1.05	97	97	97	96
P.D.	1.25*	1.13*	1.01	1.05	99	99	98	99
R.H.	1.26*	1.21*	1.16*	1.03	97	97	97	96
R.S.	1.89	1.06	1.34	0.91	95	93	98	97

Note—Standard deviation in the mean slopes was about 0.2, in the goodness-of-fit values about 0.02. The scatter plots were constructed for couples of conditions which on the x- and y-axes either differed only in lighting condition or in disparity condition. An asterisk in the slopes indicates a significant deviation from 1.0.

and the stereo cue. If this is true, one expects that the relief can be approximated reasonably well with a linear model. This is a purely phenomenological analysis, independent of any mechanistic model. Of course, the parameter values themselves (stereo base and angular position of the light source) are quite arbitrary, but at least their origins are ecologically significant: For a very small stereo base, the disparity cue vanished; for almost frontal illumination, the shading cue became very weak. On the basis of such very general considerations, one would expect a principal component analysis (e.g., Mardia, Kent, & Bibby, 1977) to yield three major components. The first component should roughly represent the “average relief” up to a depth scaling factor and can be said to measure the degree to which “shape constancy” applies. Notice that we—slightly deviating from common usage—consider two reliefs that differ only in amplitude to represent the “same shape” here. This is indeed reasonable, in view of the inherent ambiguities of the cues. One would expect the second and third components to reflect the influence of parameter variations. Higher order components then represent the influence of nonlinearities and noise. Notice that nonlinearities were certainly expected, since a linear model is only a first approximation. Table 2 shows the variance accounted for by the first three components.

As can be seen in Table 2, the first component accounted for about 94% of the variance. Thus *shape constancy* was the major single component of the response. However, the residual relief variations were both significant and had systematic regional patterns. At least two more principal components were needed to explain the systematic part of the variance. The remainder (roughly 2% of the variance) was not significant and failed to show systematic spatial variation but rather appeared noisy with only some indications of a spatial pattern.

Thus, the analysis did not contradict our assumption of a linear model. If a linear model applies, the effect of parametric variations of the cues should be revealed in projections upon the plane spanned by the second and third principal component (the 2–3 plane). In the remainder of this section, we concentrate on this plane. This allows us to disregard the overwhelming effect of shape

constancy (which turns up in the first dimension) as well as the scatter in the data (which turns up in the dimensions higher than the third). In doing this, we also discard the influence of nonlinearities, which appears to be only slight.

One expects the projections on the 2–3 plane to have the general structure suggested in Figure 7. Due to the fact that the actual parameter values are rather arbitrary and (certainly, in the case of the illumination) do not allow us to enforce an a priori metric, we decided to do a purely ordinal analysis. We simply fit a linear model to the variation in the 2–3 plane (again, this is pure phenomenology and implies no specific mechanistic model, only continuity) and considered the order of the conditions in the major directions of the 2–3 plane.

The linear fits were successful, except for the patterns for Subject B.Z. and R.S. for stimulus B, which were degenerate. Notice that the assumption of linearity implies that the order of the projections in the 2–3 plane reflects the order of increase of the parameter values. Note that it is only the order that counts; arbitrary reflections and rotations or shears are to be disregarded. For a random pattern, the probability of two violations of order is about 30%, hence we considered cases with two or more violations to be nonsignificant. This occurred for Subject R.S., for whom no single pattern was significant. Thus, for this

Table 2
The Cumulative Percentage of Variance Accounted
for by the First Three Principal Components

Subject	Principal Component	Variance Accounted for in Percentage	
		Stimulus A	Stimulus B
B.Z.	1	94.0	93.5
	2	96.4	96.5
	3	98.1	97.8
P.D.	1	96.7	98.4
	2	98.2	99.1
	3	99.1	99.4
R.H.	1	94.0	92.8
	2	96.1	96.5
	3	97.9	98.0
R.S.	1	91.6	94.9
	2	96.1	97.6
	3	97.8	98.3

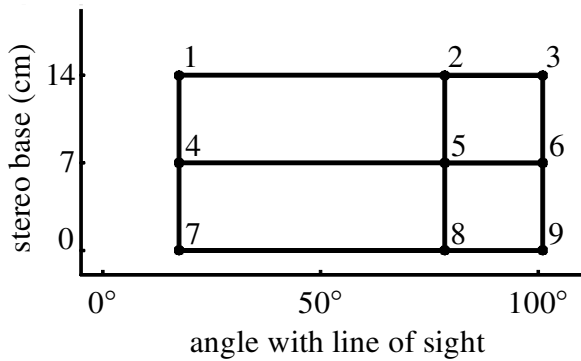


Figure 7. Parameter plot of the conditions. On the x-axis, the lighting condition is displayed as the angle with the line of sight at which the light source was positioned. On the y-axis the stereo base used is displayed in centimeters. Numbers 1–9 stand for the conditions. To reveal the pattern more clearly, the conditions are connected with lines.

subject, all systematic effects were explained by shape constancy.

For the other subjects, only 1 out of 12 patterns (the total number) was insignificant (the aforementioned case of Subject B.Z. for stimulus B). In most cases, there was one violation, and in 3 out of 12 cases there was none.

The directions of major variation in the 2–3 plane represent the influence on the relief (or the deviation from shape constancy) due to a pure disparity or illumination variation. We show these variations in Figure 8. Except for the cases of Subjects B.Z. and R.S. with stimulus B, it is apparent that the disparity cue mainly affected the global hills and curves in the vertical direction: Compared with the original shape, the bottom and the upper parts were affected so that they looked more elongated (or flattened). The illumination cue mainly caused deformations in the horizontal direction, which is indicated by the fact that, in the two bottom rows of Figure 8, mainly vertical lines can be seen. The consistency of these effects over observers is evident, even for Subject R.S. For both cue variations major changes affected the scapular and pelvic areas, for stimulus A a twist in the lumbar area.

DISCUSSION

An experiment was performed on the combination of the shading and the disparity cue. The cues were varied parametrically in a real scene, of which photographs were taken. We presented subjects with photos on a computer screen. The subjects performed local attitude settings, from which global depth maps were derived. These depth maps were analyzed with a principal components analysis. It turned out that only three components accounted for at least 97.8% of the variance in the data. As in similar experiments (see, e.g., Koenderink et al., 1996a), the first component accounted for the effect, which in literature is known as *shape constancy*. The projections of the depth maps on the second and third principal components were calculated. It was possible to isolate the ef-

fects for both the disparity and the shading cue variation as linear combinations of the second and third principal components. Apparently, these isolated effects were (globally) independent of one another: We found that the global ordinal relations between clusters in the 2–3 plane were in the expected order. A violation of this would have indicated interactions. In conclusion, we found that the effects of the cues combined linearly: The effects of one cue could be added linearly to the effects of the other. This conclusion is in agreement with the cue combination model of Landy et al. (1995). To our knowledge, however, this has never been confirmed for the shading and the disparity cue in a realistic setting. Furthermore, the individual effects are also in agreement with the literature, and, in Figure 8, it can be seen that these individual effects are very similar for each of the subjects. We think it important to mention here that this strengthens the main conclusion a great deal and is of general interest. Next, we focus on the individual effects.

From our review of the literature, we expected global linear scaling effects for the disparity cue variation. In the analysis of the scatter plots, we found that only for 2 subjects did disparity cue variation systematically lead to slopes slightly larger than 1.0. (Note that, in the scatter plots on the y-axis, depth values were portrayed with a higher stereo base than on the x-axis.) This indicates that we found only a small global scaling effect. Given the immediate impression of the stimuli, it is perhaps surprising that we did not find a larger scaling effect (e.g., see Figure 1 and compare the impression of monocularly viewing one photo of a stereo pair to the binocularly fused impression).

The plots that show the deformations caused by the different cues, as depicted in Figure 8, look similar over all subjects. Therefore we could also study a further effect of the disparity cue variation. In the two top rows of Figure 8, it can be seen that the constructed objects were affected so that the upper and bottom parts appeared more elongated in depth towards the viewer, whereas the middle parts appeared more elongated away from the viewer. Thus, global curvature was affected in the vertical direction but not in the horizontal direction. This is true for stimulus A for Subjects B.Z., R.H., and P.D., and less apparent for R.S., and also for stimulus B for Subjects P.D. and R.H. We investigated more thoroughly how this worked. It was found that especially the transition between the back and the buttocks became more pronounced with stereo base 14 cm. With stereo base 0 cm the back was slanted backwards, but with stereo base 14 cm the back became more vertical, whereas the upper part of the buttocks became more horizontal. So the third principal component influenced the angle the buttocks made to the back. This is not some kind of global scaling effect; if it were, the back would have become slanted more backward instead of less. In Figure 4, this effect can be seen. Also, locally, the shoulder blades were affected by the third principal component. In summary, we found that subjects' judgments were globally influenced in the vertical direction, in a nonlinear way.

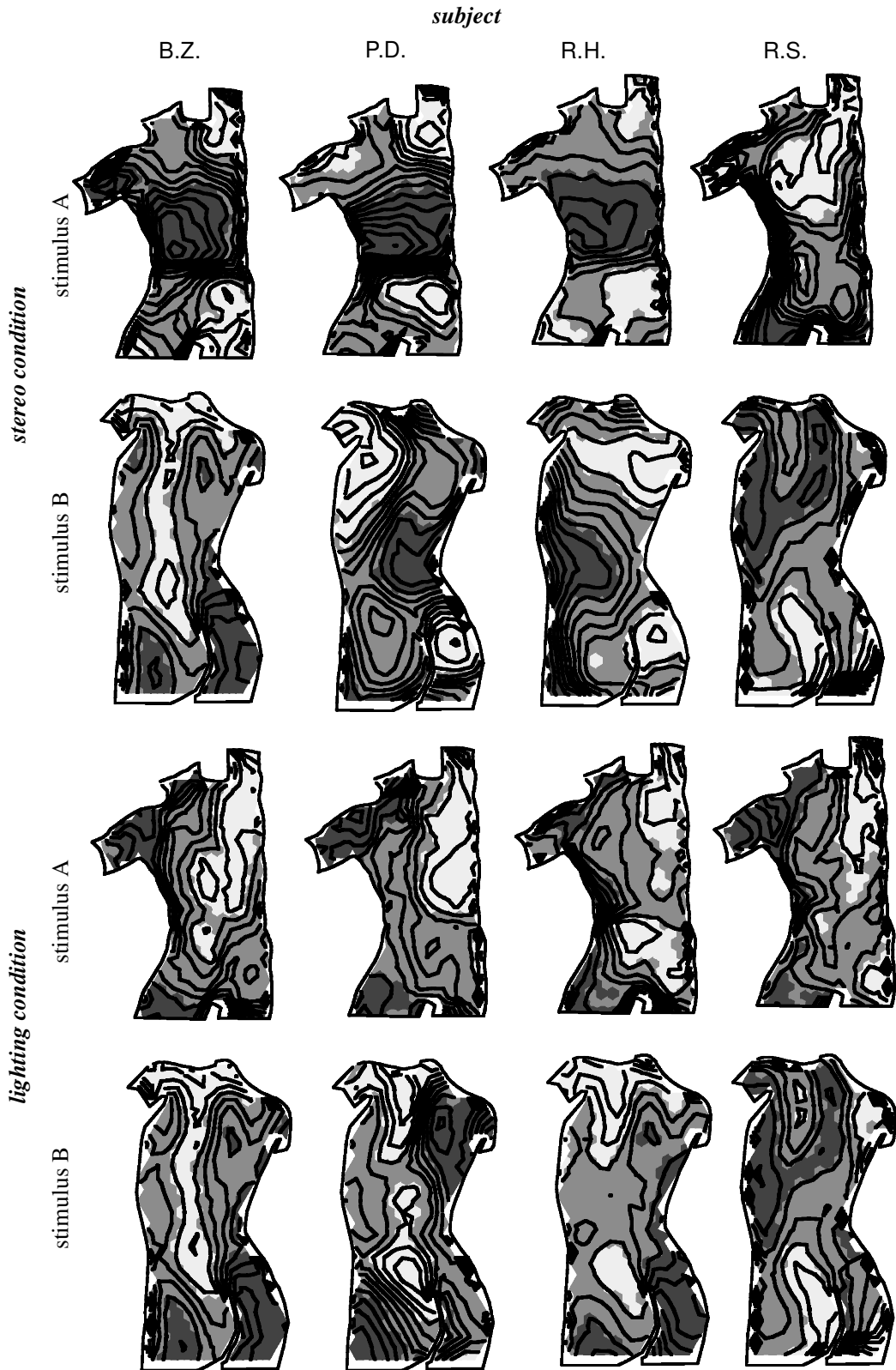


Figure 8. Deformations of the depth judgments caused by variation of the cues. The lightest and the darkest shades of gray depict the lower and the upper quartiles of the depth, respectively, the rest is displayed in medium gray. Each plot also shows 15 altitude curves. These effects were isolated as a linear combination of the second and third principal components. The 4 subjects are depicted in the columns. The two top and the two bottom rows are the isolated deformations caused by the disparity and shading cue variation, respectively.

Next, we focus on the effects of the shading cue variation. In Figure 8, it can be seen that the plots that describe the deformations caused by the shading cue variation look very similar for all subjects. Mainly vertical lines can be seen in both stimuli. This means that the effects took place mainly in the horizontal direction. Remember that the light source was also moved only in the horizontal direction. We conclude that there is a connection between the direction of the movement of the light source and the direction in which effects take place. Koenderink et al. (1996a) deduced the expectation that shading effects will be highly correlated with the component of the depth gradient in the direction of the light source. This expectation was confirmed in their experiments. So, if the light source is moved in one direction, the judged depth gradient should be altered in the same direction. Therefore, our results are also in agreement with that expectation.

There is another similarity in all deformations for stimulus A (Figure 8)—namely, that the upper part of the body is twisted with respect to the rest of the body. This can also be seen in the example depicted in Figure 5. Todd et al. (1996) also found this twist effect when varying the lighting condition for 1 subject. Again, we looked more thoroughly at all constructions and found that twists occurred systematically with the two illumination conditions in which the light source was not positioned in the line of sight. In these conditions, the right upper part of the object became twisted towards the light source, or, in other words, “brighter” (i.e., the right shoulder blade showed a highlight in these two cases) was judged as “nearer.” This is an effect that was also previously found by Koenderink et al. (1996a). However, the effect of brighter being judged as nearer did not occur for the bottom part of stimulus A, which also showed a highlight.

One comment should be made about the conclusion stated above. The twist did not occur in *all* circumstances, and, in some cases, it was barely visible. However, the fact that it showed up in all four principal components analyses for stimulus A indicated that indeed it was a systematic consequence of the variation of the shading cue.

In conclusion, in our experiment, we found individual effects for both disparity and shading cue variation, which is in agreement with the literature, and we also found that the effects of varying both the disparity and the shading cue combined linearly.

REFERENCES

- BLAKE, A., ZISSERMAN, A., & KNOWLES, G. (1985). Surface descriptions from stereo and shading. *Image & Vision Computing*, **3**, 183-191.
- BLAKEMORE, C. (1969). Binocular depth discrimination and the nontemporal division. *Journal of Physiology*, **205**, 471-497.
- BUCKLEY, D., & FRISBY, J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Research*, **33**, 919-933.
- 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.
- COLLEWIJN, H., & ERKELENS, C. J. (1990). Binocular eye movements and the perception of depth. In E. Knowler (Ed.), *Eye movements and their role in visual and cognitive processes*. (Reviews of Oculomotor Research, Vol. 4, pp. 213-261). Amsterdam: Elsevier.
- ERENS, R. G. F., KAPPERS, A. M. L., & KOENDERINK, J. J. (1993). Perception of local shape from shading. *Perception & Psychophysics*, **54**, 145-156.
- HOWARD, I. P., & ROGERS, B. J. (1995). *Binocular vision and stereopsis*. Oxford: Oxford University Press.
- JOHNSTON, E. B., CUMMING, B. G., & PARKER, A. J. (1993). Integration of depth modules: Stereopsis and texture. *Vision Research*, **33**, 813-826.
- KOENDERINK, J. J., VAN DOORN, A. J., CHRISTOU, C., & LAPPIN, J. S. (1996a). Perturbation study of shading in pictures. *Perception*, **25**, 1009-1026.
- KOENDERINK, J. J., VAN DOORN, A. J., CHRISTOU, C., & LAPPIN, J. S. (1996b). Shape constancy in pictorial relief. *Perception*, **25**, 155-164.
- KOENDERINK, J. J., VAN DOORN, A. J., & KAPPERS, A. M. L. (1992). Surface perception in pictures. *Perception & Psychophysics*, **52**, 487-496.
- KOENDERINK, J. J., VAN DOORN, A. J., & KAPPERS, A. M. L. (1995). Depth relief. *Perception*, **24**, 115-126.
- LANDY, M. S., MALONEY, L. T., JOHNSTON, E. B., & YOUNG, M. (1995). Measurement and modelling of depth cue combination: In defense of weak fusion. *Vision Research*, **35**, 389-412.
- LANDY, M. S., MALONEY, L. T., & YOUNG, M. (1991). Psychophysical estimation of the human depth combination rule. In P. S. Schenker (Ed.), *Sensor fusion III: 3-D perception and recognition* (Proceedings of the SPIE, 1383, pp. 247-254).
- MARDIA, K. V., KENT, J. T., & BIBBY, J. M. (1977). *Multivariate analysis*. London: Academic Press.
- MARR, D. (1982). *Vision*. San Francisco: Freeman.
- NORMAN, J. F., TODD, J. T., & PHILLIPS, F. (1995). The perception of surface orientation from multiple sources of optical information. *Perception & Psychophysics*, **57**, 629-636.
- RAMACHANDRAN, V. S. (1988, August). Perceiving shape from shading. *Scientific American*, **331**, 133-166.
- STEVENS, K. A., & BROOKES, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research*, **28**, 371-386.
- TNO (1972). *TNO test for stereoscopic vision* (8th ed.). Utrecht: Laméris Instrumenten B.V., Institute for Perception.
- TODD, J. T., KOENDERINK, J. J., VAN DOORN, A. J., & KAPPERS, A. M. L. (1996). Effects of changing viewing conditions on the perceived structure of smoothly curved surfaces. *Journal of Experimental Psychology*, **22**, 695-706.
- TODD, J. T., NORMAN, J. F., KOENDERINK, J. J., & KAPPERS, A. M. L. (1997). Effects of texture, illumination, and surface reflectance on stereoscopic shape perception. *Perception*, **26**, 807-822.
- YOUNG, M. J., LANDY, M. S., & MALONEY, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research*, **33**, 2685-2696.

NOTE

1. This is the square of the length of the longest semi axis of the covariance ellipse divided by the sum of the squares of the lengths of both semi axes of the covariance ellipse (Mardia et al., 1977).

(Manuscript received December 31, 1998;
revision accepted for publication October 25, 2000.)