Journal of Experimental Psychology:
Human Perception and Performance
1996, Vol. 22, No. 4, 869-878

0096-1523/96/$3.00

Surface Range and Attitude Probing in Stereoscopically Presented Dynamic Scenes

Jan J. Koenderink and Astrid M. L. Kappers
Universiteit Utrecht

James T. Todd, J. Farley Norman, and Flip Phillips
Ohio State University

The authors report on different methods to probe the structure of visually perceived surfaces in 3 dimensions. The surfaces are specified by patterns of shading with Lambertian and specular components, which deform over time and over stereoscopic views. Five observers performed 2 probe tasks, 1 involving the adjustment of a punctate probe so as to be on the apparent surface and the other involving the adjustment of a small gauge figure that indicates surface attitude. The authors found that these rather different methods yielded essentially identical depth maps up to linear transformations and that the observers all deviate slightly from veridicality in basically identical ways. The nature of this deviation appears to be correlated with the rough topography of the specularities.

Many species of animals can exploit optical structure to determine the properties of visible surfaces. Both material and geometrical properties are—at least in principle—available from the functional dependence of the radiance at a vantage point with the layout of the scene; the physical surface properties of surfaces; the motion of the observer; and the spatial, directional, and spectral properties of the sources of radiation. Several physiological mechanisms that extract such data are known to be common to species as different as man and flies. The major part of the mechanisms that constitute human visual competences is presently still unknown. One reason for this is that it has been difficult to devise methods that quantify perceived structure in sufficient detail.

The major cues available to our observers in the present experiment include the optical deformations over time and stereoscopic views of the occluding contour, Lambertian shading, and specular reflection. Observers report that these cues yield a compelling impression of a surface bounding a solid shape. This is perhaps remarkable in view of the fact that the observers are being prevented from using a variety of cues that are abundantly available in natural scenes and that almost certainly are exploited by the visual system in our daily vision. Among the more important absent cues are surface texture, cast shadows, and singularities of the occluding contour. Singularities of the contour specify depth order of surface patches at the contour, and cast shadows define depth order of spatially noncontiguous objects, whereas texture acts as a carrier for optic flow and binocular stereo and also may indicate surface slant through texture gradients. These absent cues are—without a doubt—often decisive in natural circumstances.

In this experiment there are no local marks on the surface that might reveal optical flow, whereas the only structure that might immediately reveal binocular disparity is the Lambertian component of the shading. However, even this Lambertian component is not immediately available because the retinal image is due to the compound effect of Lambertian shading and highlights. Both the contour and the specularities are different in both eyes, but their disparity does not coincide with any location on the surface. In fact, there is not really such a thing as "binocular correspondence" for these entities in the first place. Likewise, the contour and specularities transform over time, but they do not define the optic flow of any surface mark. Again, there is not really such a thing as "spatiotemporal correspondence."

For several of these sources of optically specified information, formal theories have been developed that enable us to estimate the potential contribution of this optical structure to the geometrical structure of the shape. However, these theories typically involve prior assumptions (like the availability of a set of correspondences) that are not applicable in the present case. For instance, some of the most thoroughly developed theories like "shape from motion" or "shape from disparity" do not apply at all because here one does not have "landmarks" in the sense of place labels based on simple structural properties of the image intensities. Though theories on shape from (Lambertian) shading do apply, this is

Jan J. Koenderink and Astrid M. L. Kappers, Helmholtz Institut, Universiteit Utrecht, Utrecht, The Netherlands; James T. Todd, J. Farley Norman, and Flip Phillips, Department of Psychology, Ohio State University.

This research was supported by the Human Frontiers Scientific Program Organization, Air Force Office of Scientific Research Grant F49620-93-1-0116, and North Atlantic Treaty Organization Exchange Grant CRG 920675.

Correspondence concerning this article should be addressed to Jan J. Koenderink, Universiteit Utrecht, Helmholtz Institut, Princetonplein 5, P.O. Box 80 000, 3508 TA Utrecht, The Netherlands. Electronic mail may be sent via Internet to j.j.koenderink@fys.ruu.nl.

probably only a weak cue in the present case. Moreover, even the Lambertian shading is not available as such but is compounded with the specularities. No comprehensive formal theories are available on, for instance, the transformations of the contour and the distribution of specularities—two of the most conspicuous optical structures in our stimuli. These transformations involving contour and specularities—though they may appear capricious at first sight—are without a doubt lawful, a lawfulness that the visual system is apparently able to exploit.

Occluding contour is a powerful static cue. It indicates local surface curvature and also specifies global volumetric properties (Koenderink, 1984). As a dynamic cue, the contour is even more informative (Cipolla & Blake, 1990). The importance of the contour as a depth cue has been described by Ernst Mach (1866/1959) and studied psychophysically by Norman and Todd (1994). The fact to be noted here is that the points of the object that are on the visual contour shift continually over the surface. Thus, there are no correspondences in the classical sense, as is required by many computational models for structure from motion.

Specular reflections indicate mainly regions of high surface curvature. Blake and Bülthoff (1991) and Mingolla and Todd (1986) have explored their significance for the visual perception of shape. The topological events from the dynamic structure of the pattern of specularities reveals the folds of the Gauß map (on the surface these folds correspond to the parabolic curves). Typically the highlights move over the surface and again do not define the correspondences required by most computational models, though the recent results of Norman, Todd, and Phillips (1995) indicate that human observers are able to obtain useful information from spatiotemporal changes in the pattern of specularities under certain circumstances.

Lambertian shading carries information on surface geometry (Horn & Brooks, 1989). Its interpretation (at least in the Euclidean sense) requires that the direction of illumination is available. Even if this information is lacking the topological structure of the field of isophotes reveals certain differential geometric singularities of the surface (Koenderink & van Doorn, 1980). In the dynamic case such a qualitative interpretation becomes especially powerful. If an observer moves relative to a fixed surface, the isophotes remain fixed with respect to the surface. Thus, the theory of the motion of three-dimensional (3D) curves is (in principle) applicable (Faugeras & Papadopoulo, 1990), and the spatiotemporal pattern of isophotes reveals the structure of parabolic curves on the surface. If, on the other hand, a surface moves relative to the sources of illumination, then this analysis no longer applies. The isophotes in that case will deform over the surface, thus destroying the pattern of point-to-point correspondence across successive views. Little is known at present of the extent to which human observers are able to make use of dynamic shading as a potential source of information about 3D shape. Recent results by Bülthoff and Mallot (1988) and Norman et al. (1995) show that human observers may use this information to some degree, though the precise details of how it is used remain to be determined.

In this article we concentrate on the perception of the geometry of surfaces of smooth solid objects. Intuitively, the result of optical sensing will be a field of local surface specifications such as the range, the surface attitude (slant and tilt of a small surface patch), or perhaps higher order local properties (e.g., curvature) as a function of visual direction. Although we are in no position to measure "perceived depth" as such, various methods can be designed that at least yield some operationalization (e.g., one may have an observer call out a number as one indicates a point on the surface). Such methods yield operationally defined depths. (In this article we use "range" as the physical variable and "depth" as its subjective correlate.) In principle there will be as many of such depth measures as one has methods, and there is no compelling reason why such depth measures should yield identical or even similar (e.g., identical up to arbitrary monotonic transformations) results. The possible concordance of various operationalizations is an empirical issue. In case it would turn out that a variety of methods yield essentially similar results, it would become unnecessary to label these "depths" by their corresponding method, and one may speak of "the" depth or (although this does not add additional meaning) "perceived depth." Additional evidence for concordant scales increases our confidence in the utility of the concept of an underlying single variable. Similar methods are used to establish distance scales in astronomy or time scales in archeology. We learn from such examples that no single operationalization need be compelling by itself when a concordance of essentially very different operationalizations can be established.

In this article we address the issue of possible concordance of one pair of essentially very different methods for determining the depth relief of smooth surfaces of solid objects. In one method the observer adjusts a punctate probe so as to be apparently on the surface. In this method the probe is constrained onto a line that is transverse to the surface. In the other method the observer adjusts the attitude of a small gauge figure so as to appear as tangent to the surface. The center of this gauge figure is constrained to the surface. In the latter case a range field is obtained by integration of the empirically obtained field of contact elements (local tangent planes). Both range fields can be compared with each other as well as with the actual range, referred to as the "ground truth."

The method of probing surface relief with a local point probe in binocular stereoscopic depth was designed by Gregory (1966; he calls it "Pandora's box" in later publications) and has been used by Bülthoff and Mallot (1988). The method of probing local surface attitude by means of a gauge figure that functions effectively as a surface contact element (Burke, 1985) was pioneered by Stevens (1983a, 1983b). The gauge figure is essentially "Tissot's indicatrix," conventionally used to describe attitude in topography (Strubecker, 1969). Koenderink, van Doorn, and Kappers (1992) have developed the figure into a powerful tool that enables one to gather extensive probings of relief.

In this article we present two experiments. In both cases we use the same stimulus and we sample at the same set of fiducial points, thus enabling subsequent comparison. In the

first experiment we probe the position in the visual direction to the fiducial points relative to a frontoparallel reference plane (all in the home pose assumed by the object as described in the Methods section). In the second experiment we probe the surface attitudes at the fiducial points.

General Method

Apparatus

The images are created and displayed on a Silicon Graphics Crimson VGXT workstation. Stereoscopic viewing hardware is used. The stereoscopic image pairs are presented using liquid crystal shuttered glasses that are synchronized with the monitor's refresh rate. The left and right views of a stereo pair are displayed at the same position on the monitor screen, but they are temporally offset. The left and right apertures of the liquid crystal spectacles shutter synchronously with the display so that each view of the pair is seen by the appropriate eye only. The cathode ray tube is refreshed at 120 Hz; thus each view of a stereoscopic half-image is updated with only half that rate, that is, at 60 Hz. The viewing distance is 76 cm, such that the 1,280 pixels wide × 1,024 pixels high display screen subtends 25.2° × 20.3° of visual angle.

Stimuli

The stimulus pattern for this experiment is the optical projection of a randomly shaped, globally convex solid object[1] (see Figure 1). This 3D object is obtained by distorting and transforming a sphere of 8.0-cm radius. The transformation is smooth so that the resulting surface is smoothly curved without discontinuities. The surface is defined by 5,120 triangular polygons. The shading and highlights are hardware interpolated within the interior of each triangular polygon, so that the resulting solid object appears smoothly curved. The underlying triangular facets are not apparent to the observer at all.

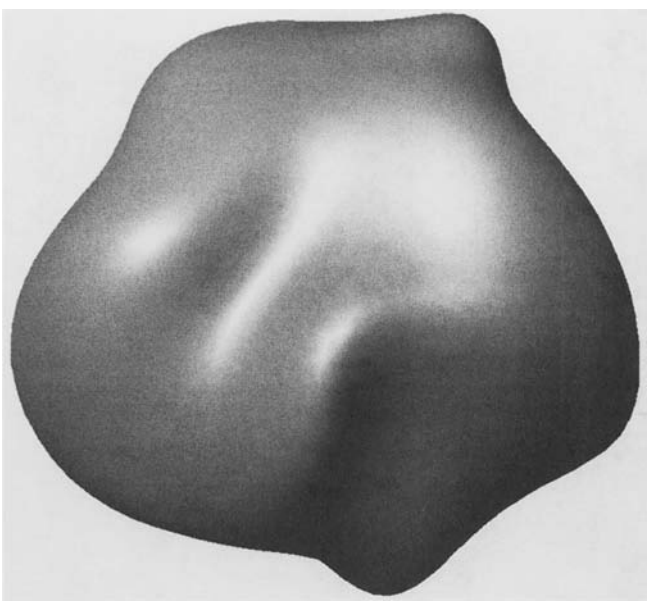

Figure 1. A monochrome rendering of the stimulus in the home position. The stimulus is presented on a RGB monitor; the Lambertian shading is in blue and the specularities are in white and are thus much more immediately evident on the screen than they are in this monochrome rendering.

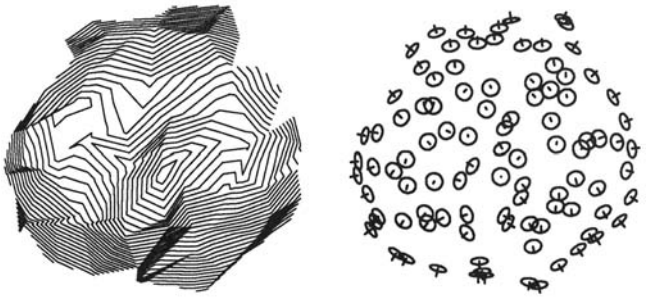

Figure 2. Representations of the ground truth interpolated from samples at the fiducial points used in the actual experiments (the sampling for the rendering on the screen is much finer). The left panel shows the congruence of equal range contours, and the right panel the field of contact elements tangent to the surface.

The object is presented as a smoothly shaded (Lambertian) blue surface combined with white specular highlights. The object appears to the observer as a "shiny blue plastic potato." The shading is generated using a standard computer graphics reflectance model (see Foley, van Dam, Feiner, & Hughes, 1990). The light source is located at infinity and is oriented obliquely at an angle of 28° with respect to the observer's line of sight. When facing the display, the illumination appears to issue forth from a source behind the observer, up and to the left (at 45° from the horizontal). The RGB values for the ambient component of the light source are (0.0, 0.0, 0.3), for the diffuse component of the reflectance (0.0, 0.0, 0.4), and for the specular component of the reflectance (0.3, 0.3, 0.3). The shininess exponent is 20.

The object oscillates between −12° and +12° from its "home" position. (see Figure 4a). The object rotates 3° at every frame transition. Therefore, each apparent motion sequence is composed of a total of nine individual frames. The individual frames are updated at a rate of 20 Hz. Each eye's view is consistent with an interpupillary distance of 6.1 cm. The projection is orthographic.

The surface was probed at 97 irregularly distributed points on the surface. These points are the same for both of the methods, thus a direct comparison is made possible.

Figure 2 graphically represents the true geometry sampled at the same points used in the probings. The left-hand figure shows the congruence of curves of equal range, sampled at equal intervals. This figure has been drawn in the pose assumed by the object at the home position. The spacing of the curves of equal range indicates the surface slant (the tighter the spacing the steeper the slope; a frontoparallel plane would have an infinite spacing), and the direction of the curves indicates the tilt of the surface. Near the occluding boundary the curves of equal range assume directions parallel to the contour (Koenderink, 1984). Though the occluding contour is the envelope of the congruence of equal range curves (i.e., has the same direction everywhere), it is not an equal range curve itself: The contour runs into depth and has distinct far and near points. Notice that the relief is dominated by two near points (range minima) separated by a saddle or pass. A major cleft

[1] "Globally convex" has to be understood in terms of scale. For instance, the earth is generally considered to be globally convex, though on the scale of the landscape one encounters local concavities.

flanked by a pair of ridges runs roughly from the lower left to the upper right of the shape. There exists an additional deep cleft at the contour of the bottom. Notice that—on a coarse scale—the shape is essentially globular.

The right-hand side of Figure 2 shows the field of tangent contact elements at the sample points. The contact elements are indicated by gauge figures that are the orthographic projections of a circular disk with a needle sticking out in the direction of the outward normal. This vividly indicates the local surface attitudes. The slant is evident from the eccentricity of the foreshortened disk as well as from the length of the foreshortened needle, whereas the direction of the major axis of the elliptical projection of the disks indicates local surface tilt. Notice that the ellipse is elongated in the direction of the curves of equal range, whereas the amount of foreshortening corresponds with the spacing of the equal range curves. The two figures thus specify essentially the same geometrical structure in different ways. Although the range and the surface attitude are different geometrical variables (range vs. surface attitude), they can be compared formally because these geometrical attributes both define a surface. This property of the stimulus is important for the interpretation of our experiments.

In Figure 3 we schematically indicate the photometric structure available to the observer. The figure shows a matrix of nine aspects, organized by rows and columns according to the following manner: The columns correspond to momentary positions in the movement of the object. The first and final columns correspond to the extreme poses, whereas the center column corresponds to the pose assumed at the home position. The rows specify various aspects of the photometric structure. The top row depicts the isophotes of the images due to both Lambertian shading and specular reflection. These components are pried apart and shown separately in the middle and bottom rows. The center row depicts the isophotes due to the Lambertian component, whereas the bottom row depicts the isophotes due to the specular component. If one attempts to fuse images column by column it becomes evident that the isophotes do carry some disparity information, though certainly not of the kind generally considered a stereoscopic rendering of the shape of the object.

Observers

Five observers (the authors) performed both tasks three times each. All have normal sight and have normal binocular stereo. As it turned out, 1 observer (F.P.) confused the inward from the outward normal in a small number of cases of the surface attitude probing task. We are convinced that this was a mistake rather than a lack of visual competence in the task. Rather than hand edit these mistakes (flipping tilt by 180°), we decided to omit the data for F.P. from the present report. A thorough check reveals that F.P.'s data are in no essential respect different from those of the other observers.

Experiment 1

Method

In the range-probing task (Figure 4b) the observer controls the position of a probe in stereoscopic space. The probe is a bright red

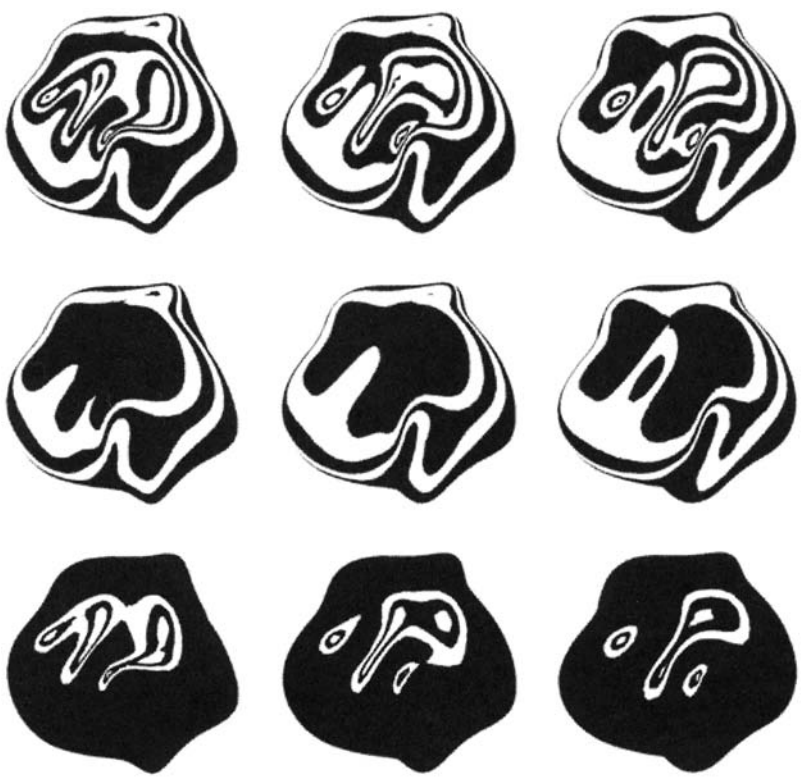

Figure 3. The illuminance distribution in the image (top row), and its components, the Lambertian shading (center row) and specularities (bottom row). The left and right columns show the poses assumed in the extreme positions of the movement, whereas the middle column displays the home position.

dot specified by binocular disparity. It undergoes the same movement as the object itself, a rotation about a vertical axis. The probe is constrained to move along a straight line transverse to the surface of the object and fixed with respect to the object. This line has been picked such that its direction coincides with the visual direction when the object arrives at its home position.

The observer uses a computer mouse device to move the probe along the line until it appears to lie on the surface of the object. The "depth" measured by this method is defined as the range of the probe when the observer is satisfied that the probe is on the surface. All points are visited in random order. A single session lasts about 30 min.

All observers find the task an easy one and rarely do they not feel confident about their settings. This is perhaps remarkable in view of the fact that the surface of the object is absolutely without texture. Relative motion (which would render the task trivial) cannot be used, not even for the smooth shading because the illuminance gradients change over time. Yet the observers do not find that they are required to set the probe "in blue air": Apparently the shading, specularities, and contour suffice to define a surface despite the fact that there is no fine-grained disparity field at all and that contour and specularities differ for both eyes, different moments in time, or both. Although at least binocular stereopsis from smooth shading exists (Bülthoff & Mallot, 1988), these effects are rather imprecise and uncertain, whereas in our experiments the object's surface appears to be rather well defined.

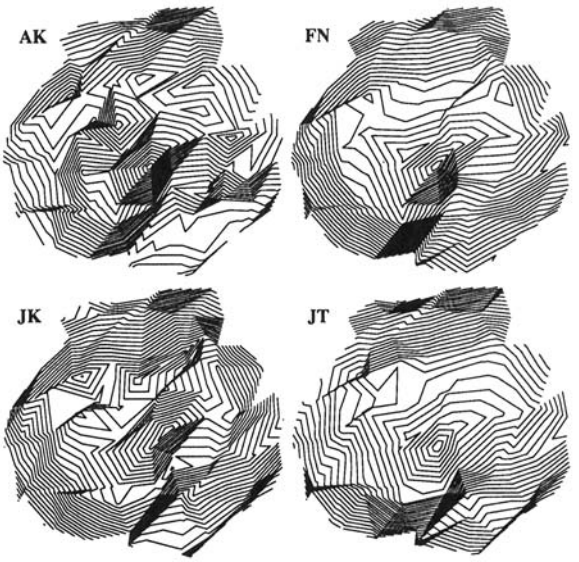

Figure 5. Results of the depth probings. Depicted are (for observers A.K., F.N., J.K., and J.T. separately) the interpolated congruences of equal depth curves based on the settings (averages over three independent sessions) at the fiducial points.

Results

In Figure 5 we present the essential results of the first experiment: The congruence of equal depth curves as interpolated from the data obtained at the fiducial points. (The reader may want to compare these results with the ground truth presented in the same format in the left panel of Figure 2.) The standard deviation in the settings of the observers amounts to about 3–6 mm (remember that the radius of the object is about 8 cm). The results in Figure 5 are the average of three independent sessions.

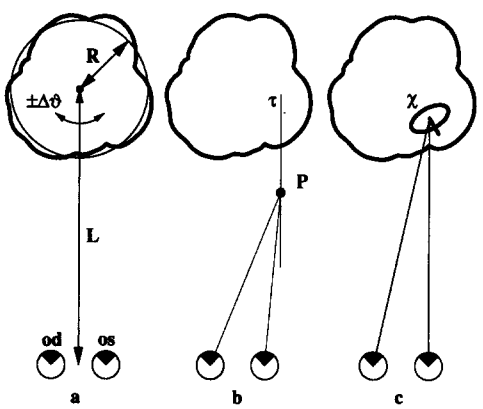

Figure 4. (a) Schematic diagram of the setup. The object rotates by $\pm\ \Delta\vartheta = \pm 12°$ about a vertical axis approximately through its center of gravity. The viewing distance **L** is 76 cm, and the radius **R** of the object is 8 cm. Viewing is binocular (eyes at od, oculus dexter, os, oculus sinister). (b) Diagram of the setup in the first experiment. The observer adjusts a punctate range probe **P** along a straight line τ until it apparently sticks to the surface. When the object is at its home position the orientation of the line τ is as indicated. (c) Diagram of the setup for the second experiment. The observer adjusts the spatial orientation of a gauge figure χ, which appears as the projection of a circle with a needle (length equal to radius of the circle) sticking out at right angles. The center of the circle is constrained to be on the surface. The observer sets the orientation of the plane of the circle until it apparently osculates the local surface.

Here we notice an unexpected phenomenon: When the probe is set to the actual surface, all observers report that it appears to hover in space in front of the perceived surface. From the intercepts of a linear regression of the data with the ground truth we find indeed that it is 11 mm behind the true surface (averaged over all observers, the range is 7.6–17.0 mm). This may be compared with the average error, which amounts (on average) to 4.1 mm, and the root mean square deviation of the residuals of the regression, which is 6.9 mm. Thus the offset is indeed significant. We have not been able to come up with a convincing explanation of these offsets though we checked the obvious possibilities.

Notice that the major features of the relief are well reproduced in the results of all 4 observers. The two near points with the cleft in between as well as the deep cleft at the bottom of the contour are evident in all of the results. Several subsidiary depth extrema are also evident, but are not or are only barely significant.

Discussion

In Table 1 we present the correlations of the observer's settings with the ground truth. We have calculated a multiple regression of the settings with a linear combination of the Cartesian coordinates of the fiducial points in the

Table 1
Affine Correlations of Depth From Range Probing

Observer	1	2	3	4	True Z
1. A.K.	—	.92	.91	.91	.87
2. F.N.		—	.95	.96	.95
3. J.K.			—	.95	.96
4. J.T.				—	.99

image plane and the range (including a constant).[2] This means that we look for the best fit of the data to the ground truth, including a depth shear and possibly a depth scaling. The correlation coefficients reported in the table are those of the settings with the best aforementioned linear combination. (We refer to these as the "affine correlations with the ground truth.") Clearly, these correlations are quite high.

Notice that the mutual correlations between the observers are also high, not surprisingly because every observer individually correlates quite well with the ground truth. In some cases (notably those involving observer A.K.) these mutual correlations even exceed that of the ground truth. Notice, however, that for every correlation a unique depth shear and scaling is applied.

Although we find that the depth settings correlate well with the ground truth (see Table 1), there appears to be an ideosyncratic linear depth scaling involved. (We found factors of 1.16 for observer A.K., 1.16 for F.N., 1.28 for J.K., and 1.09 for J.T. The deviations from 1.0 are significant for all observers. We have not yet been able to find a correlation with visual characteristics of the observers.) The affine correlation skews the "frontal plane" used in the correlation. The positions needed for optimum correlation are as follows: Observer A.K.: slant 3°, tilt −40°; F.N.: slant 4°, tilt −34°; J.K.: slant 1.5°, tilt −11°; J.T.: slant 1°, tilt 10°. Thus the slants are really small, whereas the tilts are in roughly similar directions (SD = 19°).

Although it is clear that the results show an encouraging similarity to the ground truth, it might be the case that observers assume a roughly globular surface centered within the outline. For instance, the congruence of equal range curves is roughly similar to a congruence of concentric circles centered about the center of gravity of the outline (see Figure 2, left panel). Such an assumption would certainly explain at least some of the structure of the response. To check this out we compare the results both with the ground truth as well as with a best-fitting paraboloid centered at the center of gravity of the outline. It turns out that there is indeed some correlation with the paraboloid (A.K.: .72, F.N.: .81, J.K.: .85, and J.T.: .83). This is to be expected because the ground truth also conforms somewhat to the paraboloid (r^2 = .84). The correlation with the ground truth is significantly higher for all observers (Table 1).

Deviations from the ground truth appear to reveal no systematic pattern. We return to this issue in the discussion of the comparison of the two methods.

Experiment 2

Method

We use a bright red circle as a gauge figure (see Figure 4c). The circle is rendered as a wire frame and thus does not occlude parts of the surface. The center of the circle is fixed on the surface, but the plane of the circle may assume arbitrary attitudes. The circle is defined monoscopically, not stereoscopically. The observer controls the attitude of the plane of the circle by means of the computer mouse. The circle is affixed to the object in the sense that if the observer is not manipulating the mouse, the circle is just carried along with the object. The task is to set the attitude of the plane of the circle in such a way that the circle appears to be "painted on the surface of the object." The "surface attitude" measured by this method is defined as the attitude of the plane of the circle when the observer is satisfied that the circle appears to lie on the surface of the object. The gauge figure task has been used before by the authors on both pictorial surfaces (Koenderink et al., 1992) and on real (3D) ones (Koenderink, van Doorn, & Kappers, 1995). However, in these experiments the surface was always well defined by its texture.

Again, the points were visited in random order. A single session took about 30 min. This experiment was performed after the conclusion of Experiment 1.

All observers found the task an easy one. This is indeed remarkable given the fact that the patch of the image that coincides with the area occupied by the gauge figure is typically of a uniform blue. Again, the observers cannot base their settings on local disparity (the gauge figure is presented monoscopically) nor, because of the absence of surface texture, on movement parallax cues. Apparently the "surface" is induced by features such as the contour, specularities, and shading at quite disparate locations.

Results

The main results of this experiment are presented in Figure 6. (The reader might want to compare these results with the ground truth presented in Figure 2, right panel.) Although these data are perhaps less easy to interpret by eye measure than the congruences of equal depth curves, with some effort the major near points, the diagonal cleft, as well as the sharp cleft at the bottom of the contour can be distinguished. (At near and far points the gauge figure appears like a circle with a center dot. Near the clefts one watches for sharp changes in the orientation of the elongated—because they have been foreshortened—gauge figures.)

We find that the attitude settings correlate well with the ground truth (see Table 2). Here we present correlations of the depth gradients, which are two-dimensional vectors in the image plane. We define the correlation $\rho_{\mathbf{ab}}$ between two sets of vectors $(\mathbf{a}_i, \mathbf{b}_i)$ in the following (obvious) manner

[2] Let the coordinates in the picture plane be (x, y), the range be r, and the depth be d. Then we set $d = ax + by + cr$ and find values of the coefficients (a, b, c) that yield the best fit of the (x, y, d) to the (x, y, r) in the sense of the least squares.

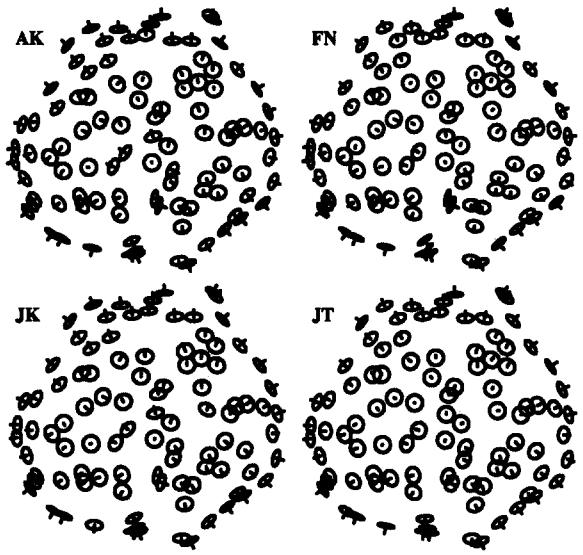

Figure 6. Results of the surface attitude probings. Depicted are (for observers A.K., F.N., J.K., and J.T. separately) the fields of surface attitudes represented by projections of a simple gauge figure. (Circular disk with needle sticking out at right angles.)

(the dot indicates the scalar or dot product):

$$\rho \mathbf{ab} = \frac{\sum_{i=1}^{N} \mathbf{a}_i \cdot \mathbf{b}_i}{\sqrt{(\sum_{i=1}^{N} \mathbf{a}_i \cdot \mathbf{a}_i)(\sum_{i=1}^{N} \mathbf{b}_i \cdot \mathbf{b}_i)}},$$

where $\mathbf{a}_i$ and $\mathbf{b}_i$ have been normalized by subtracting their means. The subtraction of the means has an effect similar to that of the affine correlation method discussed in the first experiment. It is evident that the correlations with the ground truth are very high.

In Tables 3 and 4 we present the correlation of the slants and the tilts. (In the case of the tilts we use a measure of circular correlation, basically just the aforementioned vector correlation applied to unit vectors in the tilt directions.) Notice that the slants correlate reasonably well and that the tilts correlate very well indeed, both for the case of the settings with the ground truth as well as for that of the observers with each other.

Table 2
Gradient Correlations From Attitude Probing

Observer	1	2	3	4	True gradient
1. A.K.	—	.93	.96	.93	.96
2. F.N.		—	.91	.88	.90
3. J.K.			—	.93	.96
4. J.T.				—	.95

Table 3
Affine Correlations of Slant From Attitude Probing

Observer	1	2	3	4	True slant
1. A.K.	—	.86	.90	.88	.92
2. F.N.		—	.82	.82	.85
3. J.K.			—	.85	.91
4. J.T.				—	.91

Table 4
Affine Correlations of Tilt From Attitude Probing

Observer	1	2	3	4	True tilt
1. A.K.	—	.92	.97	.91	.97
2. F.N.		—	.94	.92	.92
3. J.K.			—	.93	.94
4. J.T.				—	.99

Again, the data correlate with the best-fitting paraboloid: Observer A.K.: .91, F.N.: .87, J.K.: .90, and J.T.: .91. Although these values are high the correlations with the true shape are significantly higher for all observers. A more sensitive test is to look at the components of the gradient, namely the slant and tilt. The tilts correlate much less well with those of the best paraboloid. We find for observer A.K. the correlation with true tilt = .97 and with paraboloid tilt = .78, for F.N. the correlation with true tilt = .92 and with paraboloid tilt = .82, for J.K. the correlation with true tilt = .94 and with paraboloid tilt = .82, and for J.T. the correlation with true tilt = .99 and with paraboloid tilt = .88. The high values for the correlations with the true tilt reveal the similarity of the data to the ground truth especially well.

Figure 7 displays scatterplots of the data from observer

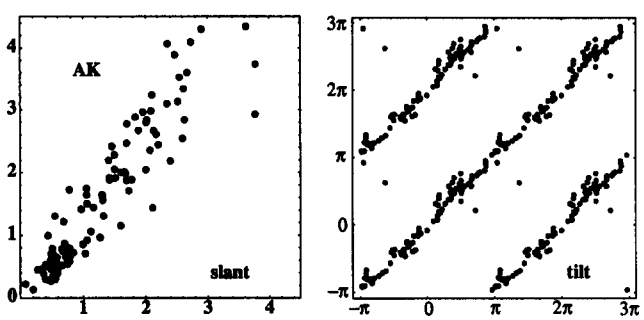

Figure 7. Scatterplots of the data of observer A.K. compared with the ground truth. The left panel shows the correlation of the slants and the right panel that of the tilts. The tilts are expressed in radians, ranging from $-\pi$ to $+\pi$. The slants are expressed in terms of the magnitude of the depth gradient, which numerically equals the tangent of the slant angle (slant angle ranges from 0° to 90°, thus its tangent from 0 to $+\infty$). The tilt domain is periodic, and we have repeated two cycles for the sake of clarity. Notice that perfect correlation would show up as a family of parallel lines at 45° with 2π spacing.

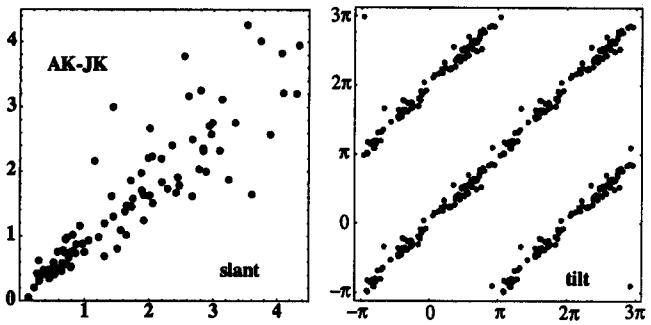

Figure 8. Scatterplots of the data of observers A.K. and J.K. The left panel shows the correlation of the slants and the right panel that of the tilts. The tilts are expressed in radians, ranging from $-\pi$ to $+\pi$. The slants are expressed in terms of the magnitude of the depth gradient, which numerically equals the tangent of the slant angle (slant angle ranges from 0° to 90°, thus its tangent from 0 to $+\infty$). The tilt domain is periodic, and we have repeated two cycles for the sake of clarity. Notice that perfect correlation would show up as a family of parallel lines at 45° with 2π spacing.

A.K. against the ground truth. The left panel shows the slants, the right panel the tilts. It is apparent that the tilts are much tighter than the slants, though both appear to be veridical on average. In Figure 8 we show scatterplots of observers A.K. and J.K. compared with each other rather than with the ground truth. It is evident that the tilts correlate especially well; observers are somewhat closer to each other than to the ground truth (Figure 7). Such a pattern is not discernible in the slants.

Discussion

The data from the second experiment are interpreted as a discrete sampling of a field of depth gradients. Such an interpretation is viable when it can be demonstrated that the field is indeed a possible gradient field, that is, that the curl of the field vanishes within the variability of the data. We have published details on this in an earlier article (Koenderink et al., 1992). Here it is sufficient to state that we essentially fit the best approximating surface to the data in a least squares sense. Such a procedure discards the curl component and thus has the effect of smoothing the data. A residual analysis reveals that the discarded components are spatially randomly distributed and that their variance is indeed fully explained by the scatter found from repeated trials. This proves that the field of surface attitude probings is consistent with samples from a smooth surface even though the samples have been obtained in (spatially) randomized order. The standard deviation in the depth values obtained in this way is 2–4 mm, which is slightly better than what we obtained for the range probe method.

When the surface has been fitted we have effectively transformed the attitude data into range data, and we may at least formally compare these values with the ground truth (Figure 2, left panel). In Table 5 we present the affine correlations. What is remarkable from these figures is that the mutual correlations of the observers tend to be larger than the correlations of the individual observers with the ground truth. This finding indicates that observers are more like each other than like the ground truth, that is, they all deviate in a similar way from the ground truth. We will return to this observation later.

As in the previous experiment, there exist ideosyncratic depth scalings (A.K., 1.39; F.N., 1.41; J.K., 1.28; and J.T., 1.25). These values are higher than in the case of range probing, and again no clear pattern can be discerned. The affine correlations skew the frontoparallel as follows: A.K.: slant = 8°, tilt = 43°; F.N.: slant = 11°, tilt = −17°; J.K.: slant = 12°, tilt = 47°; J.T.: slant = 5°, tilt = 27°. Again, the tilts appear to cluster in a systematic way (standard deviation = 25°); the slants are larger than in the case of range probing.

Table 5
Affine Correlations of Depth From Attitude Probing

Observer	1	2	3	4	True Z
1. A.K.	—	.98	.99	.95	.96
2. F.N.		—	.98	.92	.93
3. J.K.			—	.96	.96
4. J.T.				—	.93

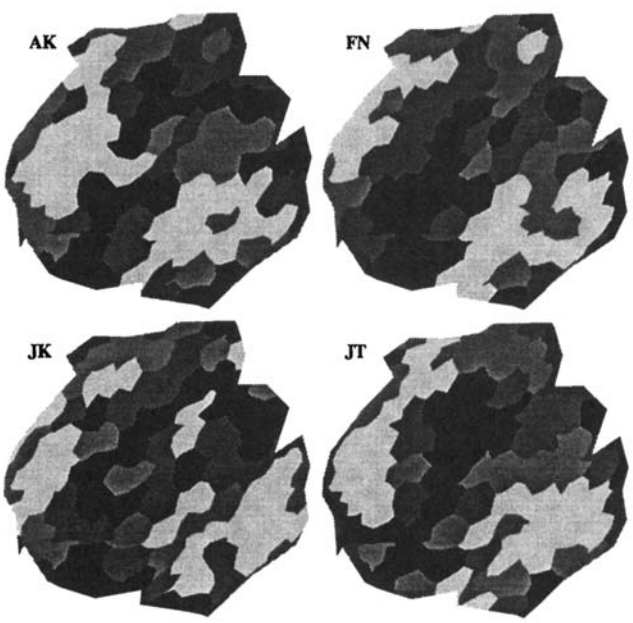

Figure 9. Residuals maps for the depth probings for observers A.K., F.N., J.K., and J.T. Notice the apparently random character. Black areas stick out toward the observer with respect to the fiducial regression surface whereas white areas recede. The range has been divided at the median and quartile values.

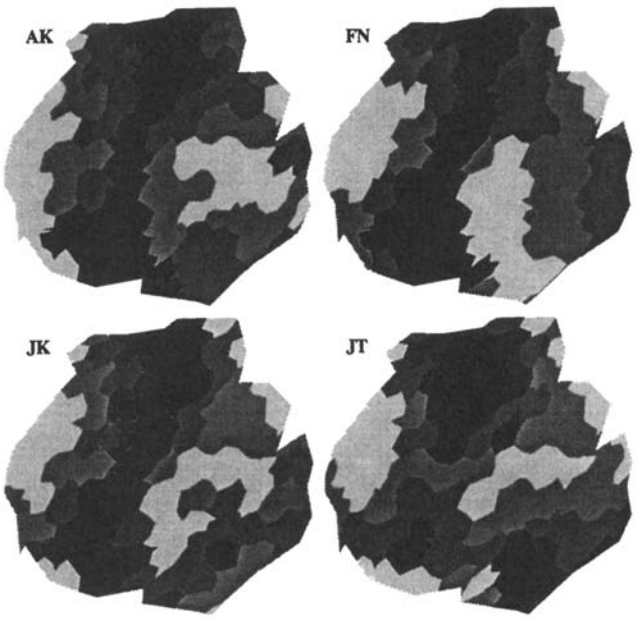

Figure 10. Residuals maps for the surface attitude probings for observers A.K., F.N., J.K., and J.T. Notice the clear pattern common to all observers: a ridge that sticks out along the bottom left to top right of the image. Black areas stick out toward the observer with respect to the fiducial regression surface whereas white areas recede. The range has been divided at the median and quartile values.

Comparison of the Range and Surface Attitude Probing

The structure in the data in both experiments is perhaps best appreciated from a study of the residuals for the affine correlation with the ground truth. We present maps of these residuals in Figures 9 (the range-probing experiment) and 10 (the surface attitude probing experiment). As anticipated, the pattern is not very clear in the first case, but rather pronounced in the second case. For the surface attitude probing the residuals all conform to a common pattern: The general area of the major cleft sticks out toward the observer relative to the best-fitting plane (basically almost the frontoparallel plane in all cases).

This is clearly demonstrated when we attempt a linear correlation of the residuals. These values are presented in Tables 6 and 7. For the case of the surface attitude probing, observers A.K., F.N., and J.K. correlate highly, and even J.T.'s correlation is significant. For the case of the range probing the correlations are much lower. (Although these are also significant the pattern is apparently different, as the two residual maps correlate only slightly as follows: for observer A.K. $r^2 = .21$, for observer F.N. $r^2 = .34$, for observer J.K. $r^2 = .23$, and for observer J.T. $r^2 = .08$.) Such a pattern would result from a nonlinear transformation of the depth values or—alternatively—from local effects due to the nonuniform distribution of cues. In the present case the major features of the luminance are indeed oriented similarly to the ridge sticking out in the residuals map (see Figure 3). On the basis of the present data we are not able to distinguish between these alternatives, although the latter alternative seems a likely one.

It is clear (at least in the case of the attitude probings) that the observers deviate in essentially similar ways from the ground truth. If the cause is to be found in the nonuniform distribution of image structure, then this suggests a powerful novel method to study the combination and interaction of multiple depth cues.

Regardless of the clear pattern in the residuals, the major results are best characterized as follows: First, both probing methods correlated well (A.K. = .86, F.N. = .92, J.K. = .94, and J.T. = .91) and yielded depth scales that were essentially related through a linear transformation (and may thus be calibrated with respect to each other). Second, observers yielded very similar results (the reliefs were related through linear transformations). Third, with both probing methods we obtained depth maps that were close to veridical.

General Discussion

We have found that both methods of probing range (one of them admittedly indirect) yield essentially identical scales (related through linear transformations). Thus, there appears to be no need to distinguish "relief from attitude probing" from "relief from range probing," as both may be calibrated to fit a single surface. This general concordance of the scales is one major result of these experiments. It enforces the idea that these very different methods may effectively probe a single underlying data structure.

It turns out that surface attitude probing yields a slightly smaller standard error for the estimated range than range probing provides, though the difference is indeed slight. Surface attitude probing provides only relief; the absolute distance cannot be retrieved from mere tangent plane orientations. Range probing can also provide absolute range

Table 6
Correlations of the Residuals From the Depth Correlations for the Range Probe

Observer	1	2	3	4
1. A.K.	—	.65	.53	.58
2. F.N.		—	.43	.53
3. J.K.			—	.41
4. J.T.				—

Table 7
Correlations of the Residuals From the Depth Correlations for the Surface Attitude Probe

Observer	1	2	3	4
1. A.K.	—	.80	.82	.57
2. F.N.		—	.83	.44
3. J.K.			—	.63
4. J.T.				—

information (distance from the fiducial frontal plane), though we find as yet unexplained offsets of more than twice the standard deviation in the individual samples.

We find that the relief obtained by means of these local probings is in the first approximation the same for all observers and that it conforms rather closely to the ground truth. Thus, these methods yield results that are very close to veridical. The main deviations from the ground truth are the sometimes pronounced depth scalings (for surface attitude probing up to 40%, for range probing up to 30%) accompanied by shears that rotate the apparent frontoparallel (by as much as 12° in the case of surface attitude probing and 4° in the case of range probing). There appears to be a systematic effect in the skewing of the frontoparallel, as the tilts are in the 37° range for surface attitude probing and 50° for range probing (two times the standard deviation; the maximum possible difference is 180°).

Although both methods yield results that are close to veridical, in the next approximation we detect relatively minor though indeed significant deviations from the ground truth, especially in the case of attitude probing. These deviations appear to be very similar for all observers and require the identification of a common cause. The nonuniform distribution of important image features is one candidate cause and, indeed, the general layout of such features appears to conform to the pattern of residuals. This is an issue that merits further attention, as it might open a new way to address problems of cue combination. However, another possible cause would be a nonlinear relation of depth to range for all observers in the case of attitude probing. In this case one needs to assume that such an effect acts similarly in all observers, thus the explanation is perhaps somewhat less parsimonious. To establish the true cause, additional experiments with different feature distributions are required.

An important conclusion from these experiments is the fact that range probing by means of a stereoscopically defined punctate probe and by means of gauge figure adjustment work very well for the optical deformations of shading and occlusion contours, even in the absence of texture and with the presence of specularities that move over the apparent surface. The surface appears well defined despite the fact that neither optic flow nor disparity may be simply interpreted for either specularities or contours, which are indeed major elements of the image. As is evident from Figure 3 the specularities move in a most capricious manner as the object rotates. Yet the surface looks clearly defined, suggesting that the visual system definitely exploits the lawfulness of these phenomena. Informal observations suggest that neither the specularities nor the contour in isolation are sufficient to induce compelling percepts of a surface bounding a solid body, though their combination is sufficient often in the absence of Lambertian shading. Recently, Norman et al. (1995) have shown that these cues indeed yield estimates of surface attitude that are on par with estimates obtained with fully textured surfaces.

References

Blake, A., & Bülthoff, H. H. (1991). Shape from specularities: Computation and psychophysics. *Philosophical Transactions of the Royal Society of London B, 331,* 237–252.

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.

Burke, W. L. (1985). *Applied differential geometry.* Cambridge, England: Cambridge University Press.

Cipolla, R., & Blake, A. (1990). The dynamic analysis of apparent contours. *Proceedings of the Third International Conference on Computer Vision,* 616–623.

Faugeras, O. D., & Papadopoulo, T. (1990). A theory of motion fields of curves. *International Journal of Computer Vision, 10,* 125–156.

Foley, J. D., van Dam, A., Feiner, S. K., & Hughes, J. F. (1990). *Computer graphics.* Reading, MA: Addison-Wesley.

Gregory, R. L. (1966). *Eye and brain.* New York: McGraw-Hill.

Horn, B. K. P., & Brooks, M. J. (Eds.). (1989). *Shape from shading.* Cambridge, MA: MIT Press.

Koenderink, J. J. (1984). What does the occluding contour tell us about solid shape? *Perception, 13,* 321–330.

Koenderink, J. J., & van Doorn, A. J. (1980). Photometric invariants related to solid shape. *Optica Acta, 27,* 981–996.

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.

Mach, E. (1959). *The analysis of sensations and the relation of the physical to the psychical* (Rev. ed.). New York: Dover. (Original work published in German in 1866)

Mingolla, E., & Todd, J. T. (1986). Perception of solid shape from shading. *Biological Cybernetics, 53,* 137–151.

Norman, J. F., & Todd, J. T. (1994). Perception of rigid motion in depth from the optical deformations of shadows and occlusion boundaries. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 343–356.

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.

Stevens, K. A. (1983a). Slant-tilt: The visual encoding of surface orientation. *Biological Cybernetics, 46,* 183–195.

Stevens, K. A. (1983b). Surface tilt (the direction of slant): A neglected psychophysical variable. *Perception & Psychophysics, 33,* 241–250.

Strubecker, K. (1969). *Differentialgeometrie.* Berlin, Germany: de Gruyter.

Received September 15, 1994
Revision received March 13, 1995
Accepted May 24, 1995 ■