Pictorial surface attitude and local depth comparisons

JAN J. KOENDERINK, ANDREA J. van DOORN, and ASTRID M. L. KAPPERS
Helmholtz Instituut, Universiteit Utrecht, Utrecht, The Netherlands

We measured local surface attitude for monocular pictorial relief and performed pairwise depth-comparison judgments on the same picture. Both measurements were subject to internal consistency checks. We found that both measurements were consistent with a relief (continuous pictorial surface) interpretation within the session-to-session scatter. We reconstructed the pictorial relief from both measurements separately, and found results that differed in detail but were quite similar in their basic structures. Formally, one expects certain geometrical identities that relate range and attitude data. Because we have independent measurements of both, we can attempt an empirical verification of such geometrical identities. Moreover, we can check whether the statistical scatter in the data indicates that, for example, the surface attitudes are derivable from a depth map or vice versa. We estimate that pairwise depth comparisons are an order of magnitude less precise than might be expected from the attitude data. Thus, the surface attitudes cannot be derived from a depth map as operationally defined by our methods, although the reverse is a possibility.

Smooth surfaces in three-dimensional Euclidean space are locally described by classical differential geometry. In case some spatial direction (e.g., the direction of gravity in the natural landscape or the viewing direction in visual space) assumes a special role, this formalism has to be replaced by the special theory of "topographic surfaces" and one speaks of "surface relief" (Liebmann, 1902/1927). Examples include topographic relief and—in case of pictures of smooth objects—pictorial relief. The mathematical description of relief was first fully developed by Rothe (1915; see also Koenderink & van Doorn 1993, 1994) in the early part of the 20th century, although much had already been guessed (often erroneously though) by mathematicians in the second half of the 19th century. In the relief formalism, the basic quantities are the *depth* and the *depth gradient*. The depth gradient is often specified by the *surface attitude*, in terms of local surface *slant* and *tilt*. Singular points are the Morse singular points (Milnor, 1963), that are the *near* and *far points* (which are local extremes of the depth) and *saddles* (or mountain passes). Singular curves are *ridges* and *courses* (named after the crests and water courses of topography) and *contours*. The contours are envelopes of the constant depth curves, and they form the boundaries of the visible parts of objects due to self-occlusion: they do not occur (at least on the large scale) in topography but are the natural boundaries of pictorial reliefs. The structure of pictorial relief as the essential element of the perception of

solid shape in general was first discussed extensively by the sculptor Hildebrand (1893) in his influential book "On the Problem of Form." Hildebrand (as many before him, notably Leonardo) saw no essential difference between relief in the "real" visual world and that of pictorial space, at least not for purely "visual," as distinct from "motor," images. (By this, Hildebrand meant vision with a stationary vantage point and limited field of view—say a viewing distance at least three times the diameter of the object.) Such general ideas provide one motivation for the present study.

In geometry, many formal identities exist between the basic quantities and their spatial derivatives—that is, such trivialities as that the gradient is the derivative of the depth. Even the very fact that the relief describes a coherent surface in the first place is expressed by simple identities: for instance, the curl of the depth gradient vanishes identically (and vice versa: any vector field with vanishing curl is the gradient of *some* surface). Such identities are of a trivial nature (e.g., the vanishing curl expresses the fact that you won't gain height if you make a tour in some landscape that returns you to your point of departure): hence, their high degree of *certainty*. Such (formal) facts may lead us to certain expectations in the study of the *perception* of surfaces, when we naively confuse perceptual entities with geometrical objects (e.g., depth with physical range). However, in the study of perception, the basic quantities *depth* and *attitude* are *operationally defined,* and hence quite *independent* quantities. Here the above-mentioned identities are not formal trivialities at all, but subject to *empirical verification*. There is no compelling reason why the curl of an empirically determined "gradient" should vanish.

In an earlier study (Koenderink, van Doorn, & Kappers, 1992), we showed that pictorial surface coherence

actually applies within the scatter inherent in repeated measurements. In the present paper, we address the question of whether pictorial surface attitude can be considered as a spatial derivative of pictorial depth within the empirical tolerances.

By a "local surface element," we mean a patch of a pictorial surface on which only minor attitude variations (as compared with the empirical scatter) occur. On such a patch, one may expect the depth difference between two points on the piece to be related to surface attitude in a simple fashion. Pairs of points spaced in a certain direction (that of the equal depth curve, which is approximately straight on the patch) should yield vanishing depth difference, whereas with points spaced in an orthogonal direction (also measured in the frontoparallel plane) one expects maximum depth differences for a given separation of the points. The latter direction is that of the gradient, or the "tilt direction." The depth difference per unit of separation is the magnitude of the gradient, or—equivalently—the tangent of the "slant," that is, the *slope* of the surface with respect to the frontoparallel plane. The expected relation is a simple linear one: $\delta z = \tan \sigma (\cos \tau \delta x + \sin \tau \delta y)$, where $(\delta x, \delta y)$ denote the Cartesian coordinates of the separation of the point pair in the frontoparallel plane (or field of view); δz, the depth difference; σ, the slant; and τ, the tilt. This equality is up to empirical verification if we design independent operationalizations of pictorial depth (difference) and attitude; also, the scatter found in repeated sessions should reflect this relation, if it applies.

In Koenderink, van Doorn, and Kappers (1992, 1994), we reported a method to probe pictorial surface (Ames, 1925; Claparède, 1904; Schlosberg, 1941) attitude; somewhat related methods have been proposed by others (e.g., Stevens, 1983a, 1983b). Depth probing, using a variety of methods, has also been attempted before (Bülthoff & Mallot, 1992; Stevens & Brookes, 1987). In the present study, we implemented a simple method to probe local depth *differences* in an independent manner (Stevens & Brookes, 1988; Todd & Reichel, 1989). These depth differences are themselves subject to an empirical verification of surface consistency: Given arbitrary triples of points, A, B, and C, one should have $\delta z_{AB} + \delta z_{BC} + \delta z_{CA} = 0$, since, on circumnavigating the closed triangular path ABC, one should not gain or lose any pictorial depth (otherwise, we would have to deal with so-called "intransitive triples" or "circular triads"; Coombs, Dawes, & Tversky, 1970). In cases in which the set of empirically determined depth differences is indeed consistent, we show how to construct the pictorial relief up to a constant depth shift. This is an interesting and novel result, as such. Then we proceed to comparing the empirical depth-difference scatter in repeated sessions from direct measurements with values as predicted (via the linear relation described above) from independent attitude measurements. Such results can throw light on the issue of whether the attitudes derive from the depths or vice versa. This may guide us in speculations of neural implementa-

tion, as to, for example, whether depth or attitude (or perhaps curvature) maps are "more primitive" entities in the human visual system.

METHOD

Stimuli

We prepared monochrome photographs of smooth objects using 35-mm film. The prints were scanned, using a Hewlett-Packard Scanjet Plus scanner, to 8-bit, about 500-pixel-high picture files. The pictures were displayed on a Macintosh 24-bit RGB monitor. This allowed us to overlay colored line graphics over the pictures. The overlay graphics could be controlled by the subject, using a tracker ball device. No special efforts were spent in trying to control the tonal rendition, although we did check to make sure that in all cases the full scale was used effectively. (We used Adobe Photoshop for tonal scale editing.) The pictures look like quality monochrome postcards.

We have set up a photographic studio that allows us to effectively control background, lighting directions, fill-in, shadow quality, perspective, and so forth. The objects were plastic replicas of artists' renderings of human torsos (both genders) commercially available for the purpose of displaying lingerie, and so forth, in fashion shops. In our experiments, the objects were textured, illuminated so as to yield powerful shading cues, and posed so as to yield important cues as to shape from the contour. The objects could be attached in a reproducible manner to a turntable that was also height adjustable so that precise control of the pose was possible. Two different pictures were used for the present studies. In this paper, we show results from a male torso in *en dos* pose (see Figure 1).
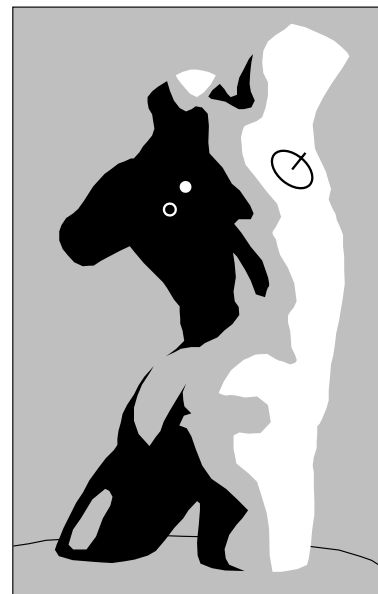
Figure 1. Posterized rendering of one of the stimuli, a photograph of a male torso *en dos*. The actual stimulus is rendered in 8-bit graytone. The white overlay suggests the indicator used for the depth-difference judgments: a pair of dots. In the experiments, the dots are distinguished by color (cyan and magenta disks) instead of structure (here disk and circle). The black overlay suggests an attitude gauge figure. The overlay figures are not to scale. The actual dimensions are discussed in the text.

We prefer this method of preparing stimuli over (otherwise far more convenient) computer graphics methods because the current renderers typically implement nonphysical processes. This applies especially to the fast hardware rendering pipelines of dedicated graphics engines. We are interested in pictorial realism rather than computer graphics renderings because we are ultimately interested in vision in the natural environment.

A patch on the pictorial plane giving rise to a contiguous patch of pictorial surface was selected by hand. This patch was automatically triangulated by a regular triangulation, which was then manually edited in order to ensure that all vertices were at locations for which the local pictorial relief was clearly defined. (Typically, we deleted a few vertices that happened to be situated on internal occluding contours.) We used triangulations of about 256 vertices. The example illustrated (the male torso in *en dos* view) had 256 vertices, 688 edges, and 433 faces. The subjects never got to see the actual triangulation during the sessions, and in fact did not need to be aware of its existence, although the probes always occurred on vertices and the subsequent analysis used the triangulation heavily.

Subjects and Viewing Conditions

The subjects were the authors, A.D. (female, slightly myopic corrected to normal), A.K. (female, emmetropic), and J.K. (male, presbyopic corrected for the viewing distance). All had normal visual acuity and good binocular stereo vision, and each used his/her preferred eye, the other eye being patched. Stimulus height was 170 mm; viewing distance was 500 mm.

Room lights were dimmed, but the monitor was visible and the subjects were fully aware of looking at a picture—not a real three-dimensional object. Due to an abundance of monocular cues, all subjects reported vivid pictorial relief, and, in fact, found it almost impossible to see the pictorial objects as merely flat.

In this paper, we show (mainly) results from Subject A.D. The results from Subjects A.K. and J.K. are, in all respects, similar though, as are the results for the other picture.

Local Depth Difference Measurements

During a session, all the *edges* of the triangulation were visited twice. During one visit, the first vertex of the edge was indicated by a dot in magenta hue (red and blue guns of the CRT fully on, green gun off; see Figure 1), the second by a dot in cyan hue (green and blue guns of the CRT fully on, red gun off); for the second visit, these colors were reversed. The visits (1,376 in a single session for the example) were made in randomized order, different for each new session. The subjects' task was to push a button indicating whether the magenta dot was closer or more distant than the cyan dot. Although the subjects tended to be puzzled when asked whether the distances were judged with respect to the vantage point or the frontoparallel plane, they did not seem to hesitate when asked to perform the task as specified. (Naive subjects often don't understand this question without further explanation, whereas they apparently need no instruction with regard to the meaning of "nearer" or "farther." We deem this remarkable and believe that the issue requires further study.) A single session took about 1 h. Subject A.D. underwent nine sessions; Subjects A.K. and J.K. underwent three sessions each. Afterwards the results were sorted with respect to the edge index and the index of the vertices per edge. The responses for the vertex 1–magenta, vertex 2–cyan and vertex 1–cyan, vertex 2–magenta pairs for each edge were classified as "first vertex in front," "second vertex in front," and "inconsistent" (or "equidistant") for each session separately.

The result for a single edge will be either consistent (first vertex always either closer or more distant) or inconsistent (first vertex once closer, once more distant). As expected, there was a clear tendency for the magenta dot to appear closer than the cyan dot, which is why we balanced the colors in the first place. From re-

peated sessions, we estimated the probability that the first vertex of an edge would be judged as being closer than, more distant than, or equally distant from the second vertex. We used Thurstone's law of comparative judgment (Torgerson, 1958) to convert these probabilities to depth inequalities, if the probability was between zero and unity, a two-sided inequality, or if either zero or one, a one-sided one. Thus we obtained a rather large set of linear inequalities (typically about 1,000 inequalities in 256 unknowns).

Attitude Measurements

During a single session, all vertices were visited three times. Each time an elliptical "gauge figure" appeared (see Figure 1), with a linepiece ending in its center, pointing into one of the directions determined by the minor axis. The subject's task was to adjust the shape of the gauge figure (using a tracker ball) to look like a circle painted on the pictorial surface, with an axle pointing into the direction of the outward surface normal. The items from a three-times-repeated list of vertices were visited in random overall order, different for each session. A single session took about 1 h. Subject A.D. underwent three sessions; Subjects A.K. and J.K. underwent one session each.

From the gauge-figure settings, the pictorial depth gradient (or attitude) was obtained by interpreting the gauge figure as a circle seen in projection. For each subject, the gradients obtained from all visits of all sessions were averaged. The size of the gauge figure had been picked for the length of the major axis to be about half the edge length of the underlying triangulation. The choice of the size of the gauge figure is, indeed, important, since the subjects perceptually *average* over surface irregularities that occur on a scale that is small with respect to the size of the gauge figure. If the size is large, we will find very smooth results, because the subject averages over an area containing several vertices of the triangulation; if the size is small, we expect very noisy results unless the picture contains evidence for a smooth surface. This is essentially the same problem that one encounters in physical measurements: the probe should be adjusted to the scale of the measurement. Smooth skin is rough under a microscope; rough bark is smooth in a blurry picture.

Numerical Methods

Several numerical methods were required in the analysis of the data.

The reconstruction of the pictorial relief from the attitude measurements and the consistency check for this have been described before (Koenderink et al., 1992).

The internal consistency check for the depth-difference measurements is fairly trivial: we merely checked all the faces of the triangulation to see whether the depth differences over the oriented boundary of a face added up to zero within the limits determined by the empirically obtained inequalities.

The reconstruction of the pictorial relief is less straightforward. *If* the system were perfectly consistent, *then* linear programming would lead to a feasible solution. However, even minor inconsistencies cannot be handled gracefully by this method. Instead, we used the much more robust method of iterated projection on convex sets. Each inequality defines a convex set in the space of all shapes—that is, the space $(z_1, \ldots, z_N)$, where z_i denotes the depth at the ith vertex, and N, the total number of vertices—to which we add the set $\Sigma_{i=1}^{N} z_i = 0$. For the example illustrated in the paper, we ended up with 1,000 halfspaces (the convex sets) and 256 unknowns. Given any configuration $(z_1, \ldots, z_N)$ and a typical convex set such as the halfspace $z_i - z_j \geqq \Delta z_{ij}$ "projection" means replacing $(z_1, \ldots, z_N)$ with $(z_1', \ldots, z_N')$ such that the new point is in the halfspace and minimizes the Euclidean distance to the initial point. If the point is already in the halfspace, we need do nothing, since the distance is already minimal. If it is not, then we set $z_k' = z_k$ for $k \neq i, j$ and $z_i' = \frac{1}{2}(z_i + z_j) + \frac{1}{2}\Delta z_{ij}$, $z_j' = \frac{1}{2}(z_i + z_j) - \frac{1}{2}\Delta z_{ij}$. For a

two-sided inequality, we may simply perform this operation twice in succession. A single projection step is thus utterly trivial to implement; however, iterating *many* such projections is *guaranteed* to lead to a solution—if one exists. We started with an arbitrary relief (e.g., the frontoparallel plane). In a single iteration step, we projected (in random order) on all convex sets. We iterated this process until either convergence occurred or we noticed that the resulting surface changed only imperceptibly. In a case of true convergence, we have found a possible solution (that is, a point in the intersection of all convex sets); otherwise, the result switches between any number of "almost" solutions. This is the major advantage of this method over linear programming. In practice, we reached (almost) stable relief solutions rather rapidly. We programmed the process in Mathematica (Wolfram, 1988), which is a fairly slow, interpreted language. The algorithm takes only a few lines of code. A few hours of iterating on a Macintosh IIfx did the job.

The numerical computation of surface relief poses few problems: *Curves of equal depth* are found via linear interpolation of the depth values over the faces of the triangulation. *Extrema* (near and far points) are found as vertices whose immediate neighbors (connected to that vertex by a single edge) are either all farther away or all closer. *Saddles* are vertices that occur as self-intersections of the constant depth curves. *Ridges* and *courses* are edges with depth gradients on both adjacent faces both facing either away from or toward that edge. Slight complications occur at the boundary or in degenerate cases, but they are simple to handle. In practice, the singular curves and points were most simply found through exhaustive search. We implemented these search processes in Mathematica (Wolfram, 1988) on a Macintosh IIfx. Finding the full relief took less than 30 min.

EXPERIMENTS

Attitude Measurements

Experiment

Subject A.D. completed a number of sessions on different days. In each of these sessions, she set an attitude gauge figure three times on all vertices of the triangulation in random order. The other subjects made three settings on all vertices in a single session. All subjects found the task very easy to do and were fast at it, acting without hesitation. When adjusted to the right attitude, the gauge figure appeared to them as if it really was on the pictorial surface.

This experiment is in all respects similar to the one described by us in an earlier paper (Koenderink et al., 1992). Figure 1 suggests what the stimulus looked like, Figure 2 shows the averaged settings of Subject A.D. Figure 3 shows the relief map as reconstructed from Subject A.D.'s data. The relief maps for the 3 subjects are very similar, indeed. Figure 4 shows the regression of the depth values for Subjects A.D. and J.K. Correlation coefficients were 0.82 for A.D.–A.K., 0.88 for A.D.–J.K., and 0.93 for A.K.–J.K.

Discussion

Data from repeated settings allowed us to estimate the scatter in the data. Figure 5 shows a scatterplot of the standard error in the component of the depth gradient in the slant direction (that is, the direction of fastest increase of depth) against the depth gradient magnitude.
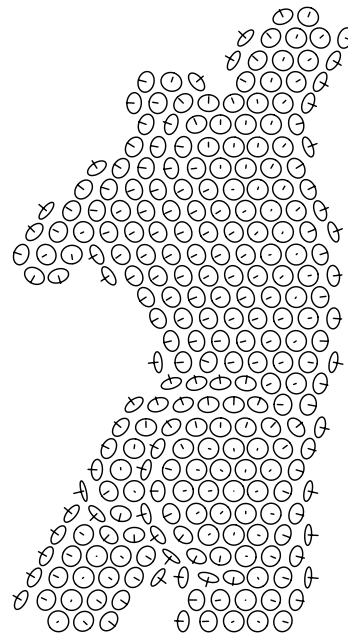

Figure 2. Gauge-figure settings for Subject A.D., averaged over three sessions held on different days. In each session, lasting about 1 h, each vertex was visited three times, so the average was over nine settings. These settings were done in randomized order, one location at a time.
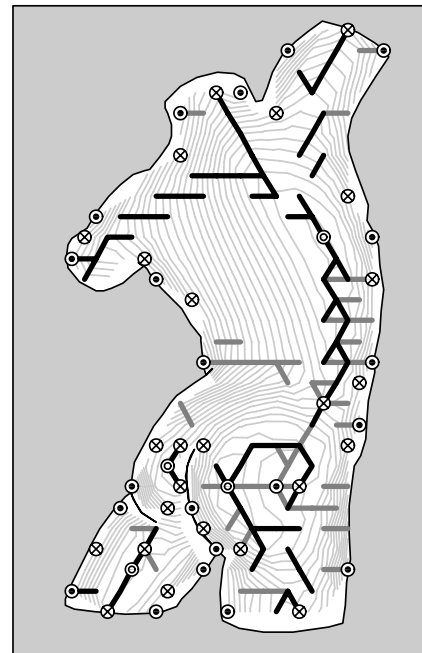

Figure 3. The full pictorial relief map calculated from Subject A.D.'s attitude settings. Drawn are the field of equal-depth curves, plus the singular points (near points, concentric circles; far points, circles with center dots; and depth saddles, circles with crosses) and curves (ridges, drawn in black; and courses, drawn in gray). Notice that the triangulation does not fully extend to the contour. The far points and saddles closest to the contour are actually on the boundary of the triangulated patch.
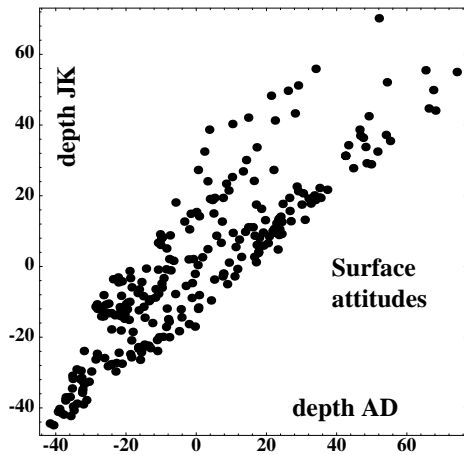
Figure 4. Scatterplot of the pictorial depths from the gauge-figure settings of Subjects A.D. and J.K. The correlation coefficient is 0.88.
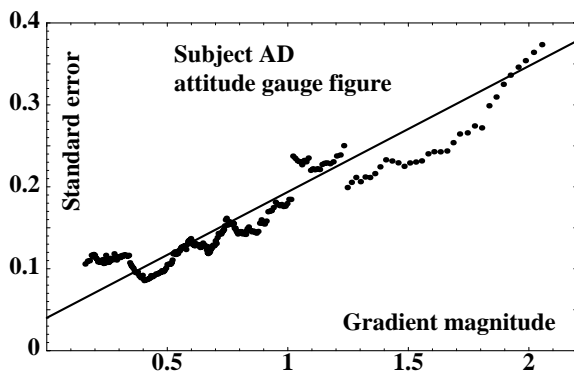

Figure 5. The standard error in the component of the gradient in the slant direction as a function of the magnitude of the depth gradient. (Remember that the magnitude of the gradient is the tangent of the slant; the slant direction is the direction of the fastest increase of depth. The unit is pictorial depth pixel per frontoparallel pixel.) The attitude gauge-figure settings are by Subject A.D. A moving median filter acting on 21-tuples of the 256 data points was used to smooth the scatterplot. Notice that this result looks much like a Weber law with a Weber fraction of about 14% and a bottom value of 0.03.

For slants that are not too large, the standard error is about 10% of the gradient magnitude with a bottom value of about 0.04 (A.D., 14%, bottom value 0.03; A.K., 4%, bottom value 0.04; J.K., 8%, bottom value 0.03).

Since the edge lengths of the triangulation measure about 20 pixels, the bottom value amounts to somewhat less than a pixel. As reported by us earlier (Koenderink et al., 1992), this scatter is highly anisotropic (see Figure 6): Almost all of the scatter is in the slant (or magnitude of the gradient), whereas the tilt can be very precise (the scatter in the direction orthogonal to the gradient is minor). It is as if the day-to-day variations were caused by small random variations in the calibration of the depth scale.

In all cases, the pictorial surfaces were consistent within the tolerance indicated by the variability of repeated set-

tings. Figure 7 shows histograms of the constraint violation over the faces (lower histogram) as well as the total depth variations over the faces (upper histogram). The constraint violations are minor as compared with the total depth variations and, indeed, amount typically to less than a pixel. (Notice that the depth domain can be measured in terms of pixels, which are, strictly speaking, defined only in the frontoparallel plane—that is, the picture plane.) The constraint violations can be fully accounted for by the day-to-day stochastic scatter in the data. It is as if the subjects were sampling from a coherent surface, rather than acting on the basis of mere local
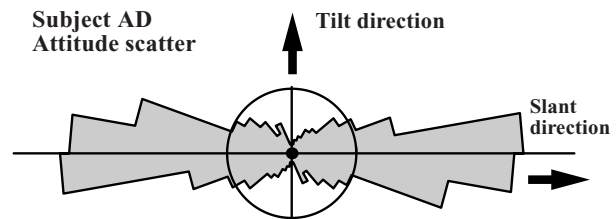

Figure 6. Polar histogram of the standard error in the depth gradient as a function of the direction with respect to the slant direction. Notice that this histogram is very anisotropic: Almost all of the scatter is in the slant direction, very little in the tilt direction. The attitude gauge-figure settings are by Subject A.D. The circle denotes the average value.
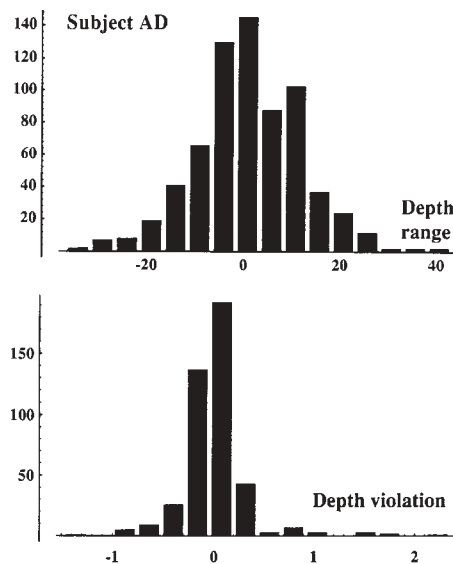

Figure 7. The upper diagram shows a histogram of the total depth variation over the faces of the triangulation. The depth variation is measured in pixels (a measure in the frontoparallel plane). The lower diagram shows a histogram of the depth violation over the faces. Notice that the violation is very minor indeed. This level of violation is fully explained by the scatter from repeated settings. The conclusion is that the attitude settings are like samples from a coherent surface. Subject A.D.

estimates. This is a remarkable result, because it appears to rule out the notion that the data can be explained by local mechanisms.

The relief maps (Figure 3 shows the relief for Subject A.D.) reveal that the subjects produced very similar results. Figure 8 shows profile views with the true profile (produced by making a photograph of the actual object turned over 90º and 270º). The empirically obtained profiles all deviate to some extent from true veridicality. The field of constant depth curves (Figure 3) reveals that

the pictorial relief is highly articulated. The singular points and curves clearly define the various anatomical features of the torso in the given pose. The relief structure is as expected: major ridge systems define the edge of the thorax turned toward the spectator, and tributary systems define the extremities (arms and legs) and the shoulder girdle and buttocks. The major course defines the edge of the lumbar region, turned away from the spectator. Major near points are found on the buttocks and the right scapula. A major internal saddle point defines the
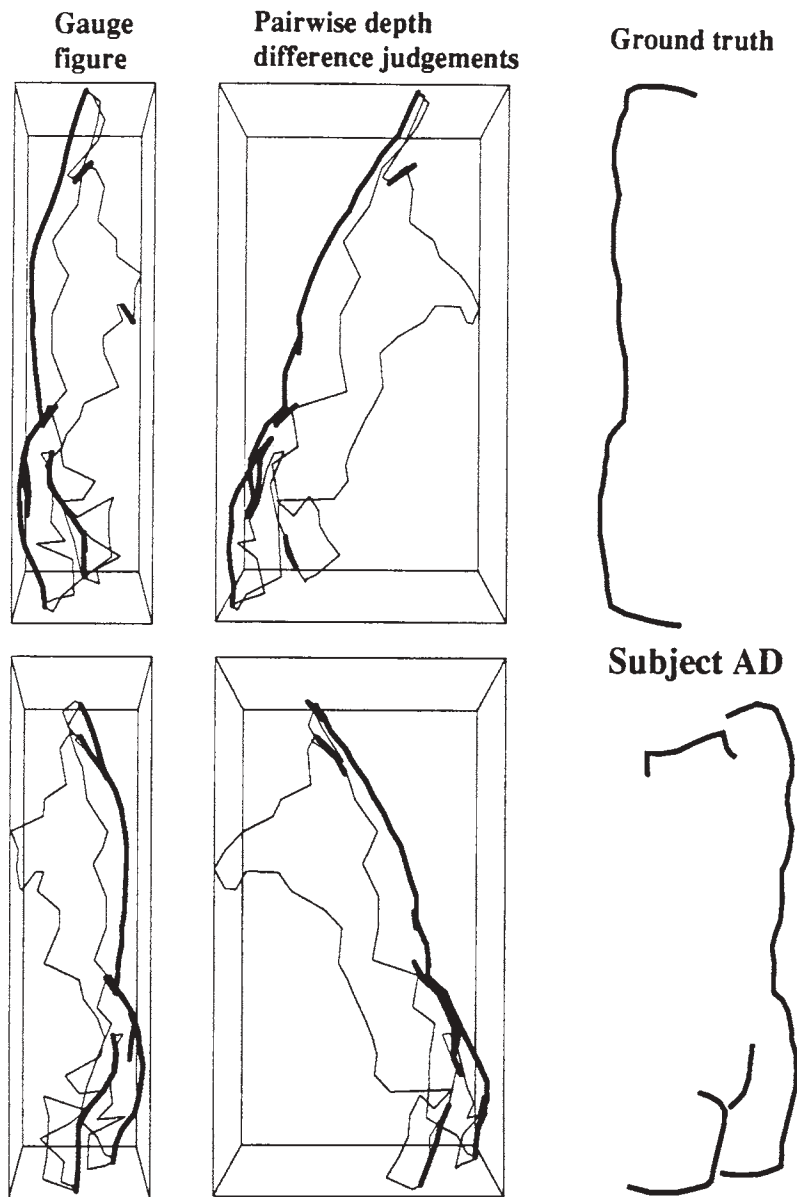
Figure 8. Right and left profile views (upper and lower rows) for the pictorial relief from the attitude gauge-figure settings (left column) and the pairwise depth-difference judgments (center column), compared with physical reality (right column). The real contours were obtained by photographing the object turned over 90º or 270º and tracing the contours with an edge detector followed by manual editing to remove slight garbage (all done in Adobe Photoshop). Notice that the depth scales cannot be compared, but that the shapes can be.

nearest point of the waist, whereas the right buttock reveals an internal far point (dimple between the gluteus maximus and the trochanter.) An extensive system of saddles and far points exists on the boundary of the triangulation, as is to be expected in a largely frontal patch.

We find that the 3 subjects seem to have utilized the monocular depth cues in essentially the same way; typically, subjects are more like each other than like physical reality, although deviations from veridicality are rarely dramatic. We will, however, not discuss the issue of veridicality in this paper.

Depth Comparisons

Experiment

The subjects completed several sessions of pairwise depth comparisons on the same pictures used for the attitude gauge-figure settings. All subjects found this task harder than the previous one, and often hesitated before coming to a decision. With a little practice (10 min or so), however, the subjects could make 400–800 judgments an hour. Although the dots *could* be seen as being on the pictorial surface, this often took something of an effort. Even if the dots appeared to "float in space," the task was still possible since they merely indicated locations in the visual field. The dots are "indicators" rather than "gauge figures." Although the subjects never complained of "mistakes" in the gauge-figure task (indeed, one can *see* whether the setting is right or not), they did report occasional mistakes in the present task, indicating a nonperceptual task component ("I meant to press for cyan but somehow pressed for magenta by mistake").

Of the single judgments, about 9 out of 10 are consistent (A.D., 87%; A.K., 90%; J.K., 87%). Of the consistent judgments, more than half (A.D., 67%; A.K., 53%; J.K., 57%) have the magenta dot in front (since we carefully balanced the design, this should be 50% if there were no effect of hue). Thus, the hue has a rather strong biasing influence, as expected. In pictures, reddish patches tend to look closer than bluish ones, an effect well known and often exploited by painters. (The influence of this bias on the final results is, of course, cancelled out with our method.) Since the subject has to be able to distinguish the dots in order to perform the task, the dots always have to differ in *some* dimension (if not hue, then geometrical structure, or temporal sequence, or something else), and any difference at least *potentially* introduces a bias. Balancing such influences in each single trial seems to be the only reasonable solution.

After averaging over all sessions, we found that for most edges the empirical probability of the first vertex to be closer than the second is somewhere between zero and unity (as opposed to being identically zero or unity). Only a small fraction of the edges (A.D., 13%; A.K., 13%; J.K., 15%) lead to an inconsistent judgment (i.e., in all sessions, the edge was judged inconsistently). Such edges were assigned 50% probability of the first vertex's being judged as being nearer than the second. Notice that such a 50% probability would also be obtained if the

subject had the first vertex (consistently) closer in half of the sessions and the second vertex closer (also consistently) in the remaining sessions. These cases might be considered to be different although they lead to the same probability of the first vertex's being closer in the final result. One might attempt to distinguish such cases, perhaps through an additional "confidence" parameter. In this study, we have not done so: We used only the final (averaged) probabilities.

The faces of the triangulation can be tested for consistency ("intransitive triples" or "circular triads"; Coombs et al., 1970) relative to the probabilities of vertex order for the edges that make up their boundaries. The consistent triangles are of one of the following types (the arrows indicate the judged depth order; we assume the boundary of the triangle to be oriented): $(\Rightarrow, \Rightarrow, \Leftarrow)$, $(\Rightarrow, \Leftarrow, \Leftrightarrow)$, $(\Rightarrow, \Leftrightarrow, \Leftrightarrow)$, and $(\Leftrightarrow, \Leftrightarrow, \Leftrightarrow)$ (all to be taken cyclically and also in reversed order, 19 types in total), whereas the inconsistent triangles are of the $(\Rightarrow, \Rightarrow, \Rightarrow)$ and $(\Rightarrow, \Rightarrow, \Leftrightarrow)$ types (also to be interpreted cyclically and also in re-
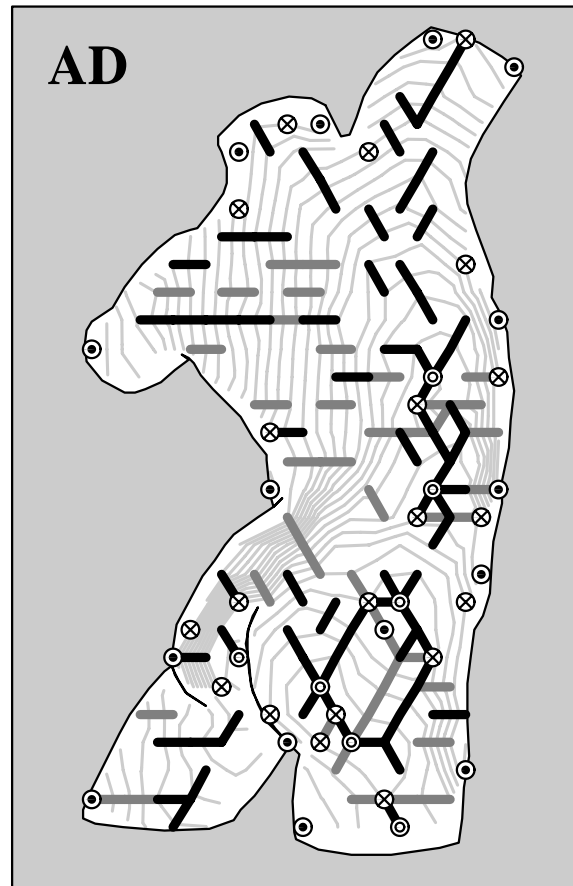

Figure 9. The full pictorial relief map calculated from Subject A.D.'s pairwise depth comparisons. Drawn are the field of equal depth curves as well as the singular points (near points, far points, and depth saddles) and curves (ridges and courses). Remarks in the caption of Figure 3 apply here as well.

versed order, 8 types in total). (This could be made a lit-
tle more precise when the estimated depth differences
are used. We used the more robust combinatorial algo-
rithm outlined here.) We found that the estimates are in-
deed fully consistent; not a single face violated the con-
dition. A search for larger cycles using graph theoretical
algorithms (Skiena, 1990) also yielded a negative result.
Thus, the pairwise depth-order judgments lead to a glob-
ally consistent set of data. Again, it is as if the subjects
based their decisions on samples from some coherent
surface rather than relying on mere local estimates.

Reconstructions of pictorial relief via the iterative
method of projection on convex sets converged on very
similar results, no matter what initial guess was taken.
(We usually set the initial estimate to the frontoparallel
plane.) Only a small number of violations remained, as
to be expected from the consistency check. The correla-
tion of the observed probabilities of judging the first
vertex of an edge as being closer and the probabilities as
predicted from the depths is over 90% (A.D., 94%; A.K.,
96%; J.K., 93%), the average discrepancy is about 1/10th
(A.D., 0.086; A.K., 0.11; J.K., 0.12), an excellent result.
The relief maps for all subjects are essentially the same;
Figure 9 shows Subject A.D.'s.

Discussion

Again, the subjects' results are very similar. Figures 9
and 10 show the relief map and oblique views of the pic-
torial surface. Figures 11 and 12 show the results of the
regression of the depth rank orders for Subject A.D. A
linear regression on the depths yields essentially the

same results, indicating that the reliefs are close to met-
rical (modulo an unknown factor) rather than merely or-
dinal. This is due to the fact that most of the probabili-
ties are between zero and unity, thus enabling us to use
Thurstone's law of comparative judgment (Torgerson,
1958). A study of the residuals reveals patterns that are
quite idiosyncratic for our subjects.

As with the pictorial relief obtained from the attitude
gauge-figure settings, the reliefs for the pairwise depth
comparison judgments are not truly veridical, although
subjects largely agree in their deviations from true ve-
ridicality.

GENERAL DISCUSSION

Comparison of Relief From the Two Methods

Although the pictorial relief surfaces as reconstructed
from pairwise depth comparisons and attitude gauge-
figure settings show differences, especially in overall
slant, in a first rough characterization they are very simi-
lar. A rank order correlation of the pictorial depths from the
gauge-figure settings and the pairwise depth-difference
judgments yields Spearman rank correlation coefficients
of 71% for Subject A.D., 75% for Subject A.K., and 90%
for Subject J.K. Because the reconstruction of the global
relief involves an integration in both cases, one should
probably ignore the conspicuous shallow overall trends.
For this reason, we preferred to use the raw local data,
rather than the reconstructed depths, for our quantitative
comparison to be presented in the next section.

Subject AD

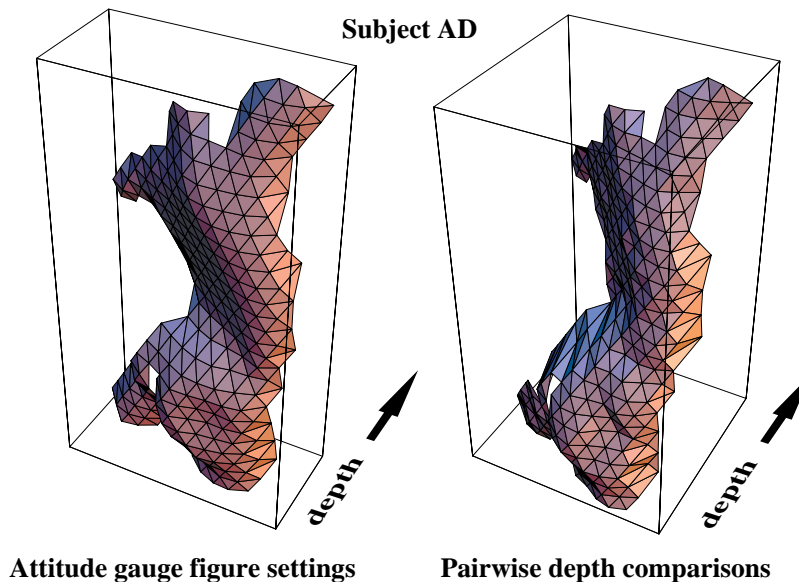**Attitude gauge figure settings**   **Pairwise depth comparisons**

**Figure 10. Oblique views of the pictorial relief surfaces as calculated from the attitude
gauge-figure adjustments and from the pairwise depth comparisons. The depth scales of
the figures are not commensurable, but it does make sense to compare the shapes. Notice
the apparent "change of pose" between these results—a torsion in the waist that twists the
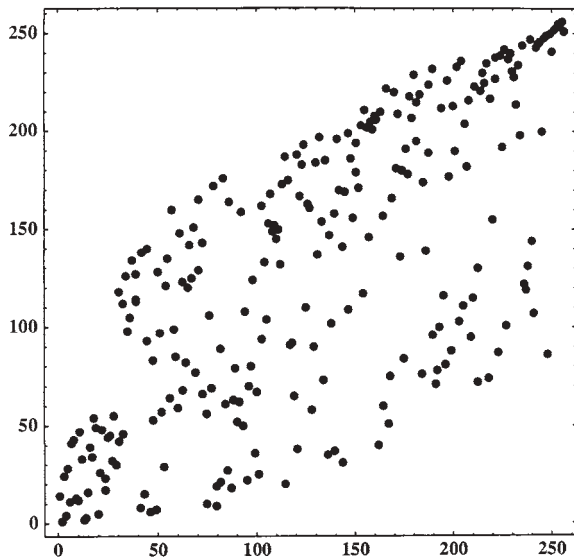thorax with respect to the pelvic region. Subject A.D.**

Figure 11. Rank correlation of pictorial depths from attitude gauge-figure settings and pairwise depth comparisons. Subject A.D.

Analysis of Scatter From Repeated Settings

We used the results presented above to compare the tolerances in the articulation of pictorial relief from the two methods of determining it. (Perhaps a better way of putting it is to say that both methods operationally *define* pictorial relief. Thus we compare the scatter in two independent pictorial reliefs due to the same stimulus but different tasks.)

The pairwise depth comparisons yield the probabilities of the first vertex of an edge being closer perceptually than the second vertex of that edge. The attitude gauge-figure adjustments allow us to numerically estimate the depth difference between the vertices of an edge in an independent manner: We calculated the value of the scalar product of the edge vector on the average of the empirically determined pictorial depth gradients at the vertices (that is, merely average edge's slope times edge length). This calculation can be shown to yield a least squares estimate of the pictorial depth difference from the attitude gauge-figure adjustments. By plotting the probabilities of the first vertex of an edge's being nearer than the second, as obtained from the pairwise depth comparisons against the depth differences from the gauge-figure settings, we obtained a "psychometric curve" (see Figure 13). There is a twist here, since we remain completely within the pictorial space, and thus the "psychometric curve" does not represent a psychophysically obtained probability as a function of a *physical parameter*. In our analysis, we did not need the physical range (as "objective correlate of the depth") at all. The procedure applies equally well if the physical configuration is not known; it could be used just as well on artistic renderings of nonexisting objects, for instance.

We found that the "psychometric curve" describes a nice ogive shape, indicating that the probability of judging the first vertex as being closer than the second increases rather slowly as a function of the depth difference between the vertices, as estimated from attitude settings. The 50% point is within a few pixels from zero, and the "standard error" is 8 pixels for Subject A.D., 4 pixels for Subject A.K., and 5 pixels for Subject J.K.; thus, the rising portion of the curve extends over a depth-difference range of about 8–16 pixels.

The uncertainty in the depth-difference estimates from the gauge-figure settings (see Figure 5) is much smaller than the width of the sloping portion of the ogive curve (the constraint violation is of subpixel dimensions, whereas the standard error amounts to somewhat less than a pixel), suggesting that the depth-difference judgments are at least an order of magnitude less precise than the depth differences found from the attitude settings. This result holds true for all subjects.

One might object the slope of the ogive curve is a measure of the reliability of a two-alternative judgment, while, in the case of the attitude data, we measure reliability as the spread in repeated adjustments of a gauge
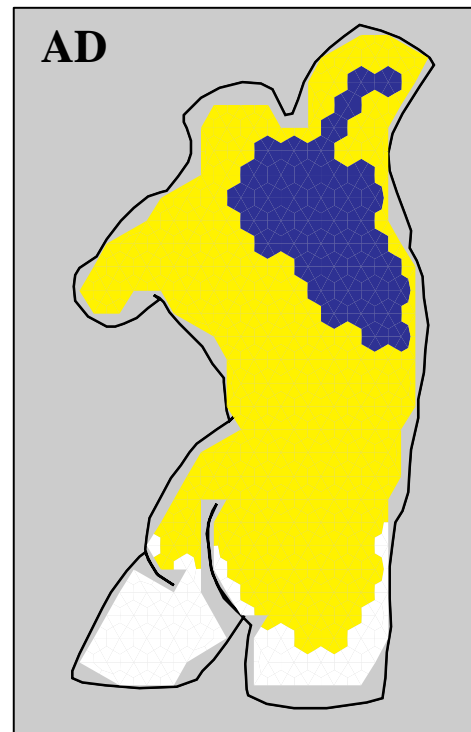

Figure 12. Residuals from the rank correlation of pictorial depths from attitude gauge-figure settings and pairwise depth comparisons. Residuals were binned into three categories: vanishing (between 0.25 and 0.75 quantiles, rendered in gray) and extreme (over 0.75 quantiles, rendered in white; under 0.25 quantiles rendered in black). Subject A.D. Notice that the deviations are clearly systematic, rather than randomly distributed. (A truly random distribution would yield a salt-and-pepper dotted pattern.)
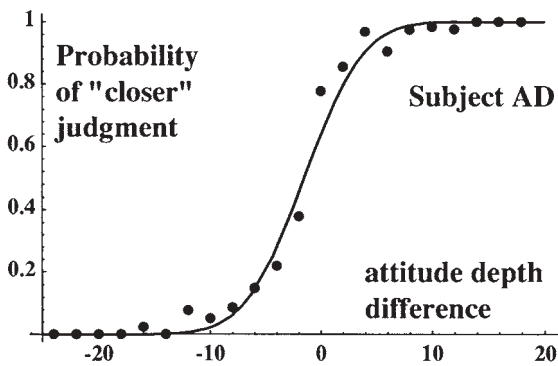
Figure 13. "Psychometric curve" for pairwise depth comparisons and attitude gauge-figure adjustments for Subject A.D. The probability of judging the first vertex of an edge as being closer than the second vertex is plotted as a function of the depth difference over the edge as calculated from the gauge-figure settings. The standard error in the latter is very small (at most, a pixel) in the region where this "psychometric curve" has a rather shallow slope. We conclude that depth (difference) from direct comparison is about an order of magnitude less precise than it is from gauge-figure settings.

figure—two quite distinct and not immediately comparable quantities. This is undoubtedly the case. Yet a pair of gauge-figure adjustments (both vertices at an edge) is an effortless task in which it takes only a few seconds to decide the issue of whether one vertex will be judged as closer than the other *on the average* in a large number of difficult pairwise comparisons. In the latter process, the subject is prone to make many incompatible judgments, whereas attitude settings yield the perceptual depth order (in that task) with hardly any variation at all.

Another objection might be that the difference in difficulty experienced by the subject in the two tasks indicates that although *perception* of depth is quite accurate, the pairwise depth-comparison task yields inaccurate results because the task is unnatural. Indeed, we cannot claim that the results will generalize over (very) different tasks. However, we fail to see how "depth" could be given an objective meaning at all *except operationally*, that is, relative to some specific task. Although the speculation that "perception might be more accurate" thus makes no sense to us, it is indeed possible that a different operationalization might be found that yields much more precise depth estimates. However, such a method is presently not available. The reconstructions of pictorial depth (Figures 9 and 10) presented here are as yet unmatched in the literature.

Conclusions

In the case of the attitude gauge-figure settings, we replicated essentially the same findings as reported earlier (Koenderink et al., 1992): One obtains a consistent field of surface contact elements, and most of the variability is in the slants, as if there were some random day-to-day dither in the calibration of the depth scale. The variability found between subjects is largely of the same

nature but typically greatly exceeds the variations for each individual subject.

In the case of the pairwise depth comparisons, we found the probabilities of judging one vertex of an edge as being closer than the other again to be fully consistent with a hypothetical "pictorial surface." We were able to reconstruct such a surface via a simple and robust iterative method. These are novel and interesting results, as such. That we found consistency is probably (at least partly) due to the fact that all pairs of vertices used in the pairwise comparison were closely spaced. In pilot experiments, we found that larger spacings introduced novel problems: The depth order is at best a partial order, certainly not a linear one. Subjects are often hard put to compare depths of points on different slopes of the pictorial relief, although the task seems doable for pairs of points that lie on a single slope. Such problems have been discussed before, both from a theoretical (Koenderink & van Doorn, 1979) and from a phenomenological (Todd & Reichel, 1989) vantage point. This appears to be a very important issue, and we will address it more extensively in a subsequent study. For the present comparison, the problem can safely be ignored though.

The reconstructed "pictorial reliefs" as operationally defined by the pairwise depth comparisons and the attitude gauge-figure settings are similar, although differences are readily apparent. Yet the depth-difference judgments do not seem to be simply related to (perhaps more primitive) surface attitude judgments, for then one would expect them to be much more reliable. Of course, it is not strictly necessary for any global "internal representation" to exist in the first place, and if there is *one* such entity there seems to be no reason why there couldn't be *several*, perhaps unrelated ones. On the other hand, purely local computations—like Marr's (1982) "2½D sketch"—will not do, for the chance that one would obtain a globally consistent result is remote. Some "top–down" process must ensure global consistency, suggesting that *surfaces* are in many respects very elementary perceptual entities (Nakayama & Shimojo, 1992). Since the subject always views the picture as a whole (masking off much of the picture would probably make the task impossible), there is no need to assume any global data structure in the head.

An issue that often comes up in discussions is whether or not the fact that subjects are undoubtedly familiar with the general shape of the human body trivializes the results. Although on the face of it a serious objection, this seems rather unlikely to us for a variety of reasons. First of all, in all experiments we have done on photographs of abstract sculpture, we have invariably found comparable results. Secondly, the subject's results are not veridical in detail, and there exist noticeable inter-subject differences. Finally, it is easy to overestimate the knowledge laymen have of the shape of human bodies (themselves of almost infinite variety) in the great variety of natural poses. It is not for nothing that those aspiring to gain proficiency as academic artists need years

of study, not just to reproduce, but literally *to see* (or gain sensitivity to) the shapes of the surface of the model.

Given the large difference in precision of the pairwise depth comparisons as compared with the attitude gauge-figure settings, we conclude that it is rather unlikely that the attitudes are derived from a pictorial depth map by differentiation. If there is any functional relation between the pictorial depths and attitudes at all (and this seems not entirely unlikely in view of the high correlation of the reliefs obtained via the two methods), it is more likely that the pairwise depth comparisons are based upon the representation of surface attitudes. This is somewhat less surprising in view of the fact that many of the monocular depth cues yield primarily *surface curvature* or *surface attitude* rather than range estimates. However, it would seem that an overall integration of multiple depth cues is perhaps most simply phrased in terms of a range map (or range modulo a factor and constant term). This is no doubt the generally accepted view of the machine vision community, and it is also implicated by various psychophysical findings (Nakayama & Shimojo, 1992), although it has been implied that surface attitude is also a primitive variable (Stevens, 1983a). That the monocular depth cues are indeed integrated into some global, coherent representation of pictorial relief is indicated by the otherwise surprising consistency of both the attitude settings and the pairwise depth comparisons.

An issue that is of paramount importance when comparing the precision by which attitudes can be obtained from depth maps is that of *scale*. Empirical depth maps cannot be differentiated, and their gradient can be estimated only by averaging over a certain local region. The size of this region determines both the resolution of the resulting gradient map and the noise amplification introduced by the process of differentiation. The averaging "regularizes" the differentiation, and any robust differencing operation represents some compromise involving the scale of the operation. In our case, we have carefully balanced the edge length used in the pairwise depth comparisons and the size of the gauge figure used with the attitude gauge-figure settings. In view of the rather large difference found in this study, the conclusion that it is highly unlikely that the attitudes are derived from the depths is quite well founded. Depth-difference judgments and attitude settings appear to be independent tasks.

REFERENCES

Ames, A., Jr. (1925). The illusion of depth from single pictures. *Journal of the Optical Society of America*, **10**, 137-148.

Bülthoff, H. H., & Mallot, H. A. (1992). Integration of depth mod- ules: Stereo and shading. *Journal of the Optical Society of America A*, **5**, 1749-1758.

Claparède, E. (1904). Stereoscopie monoculaire paradoxale. *Annales d'Oculistique*, **132**, 465-466.

Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Hildebrand, A. (1893). *Das Problem der Form in der bildenden Kunst* [The problem of form in the visual arts]. Strasbourg.

Koenderink, J. J., & van Doorn, A. J. (1979). The structure of two-dimensional scalar fields. *Biological Cybernetics*, **33**, 151-158.

Koenderink, J. J., & van Doorn, A. J. (1993). Local features of smooth shapes: Ridges and courses. In B. C. Vemuri (Ed.), *Geometric methods in computer vision II* (SPIE, Vol. 2031, pp. 2-12).

Koenderink, J. J., & van Doorn, A. J. (1994). Two-plus-one-dimensional differential geometry. *Pattern Recognition Letters*, **15**, 439-443.

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. (1994). On so called paradoxical monocular stereoscopy. *Perception*, **23**, 583-594.

Liebmann, H. (1927). Geometrische Theorie der Differentialgleichungen: Die topographischen Kurven [Geometrical theory of differential equations: The topographical curves]. In W. Fr. Meyer & H. Mohrmann (Eds.), *Encyklopädie der Mathematischen Wissenschaften mit Einschluss ihrer Anwendungen* (Vol. 3, 3rd part). Leipzig: B. G. Teubner. (Original work published in 1902)

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.

Milnor, J. W. (1963). *Morse theory*. Princeton, NJ: Princeton University Press.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, **257**, 1357-1363.

Rothe, R. (1915). Zum Problem des Talwegs [On the valley path problem]. *Sitzungsberichten der Berliner Mathemathischen Gesellschaft*, **14**, 51-69.

Schlosberg, H. (1941). Stereoscopic depth from single pictures. *American Journal of Psychology*, **54**, 601-605.

Skiena, S. (1990). *Implementing discrete mathematics*. Redwood City, CA: Addison-Wesley.

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.

Stevens, K. A., & Brookes, A. (1987). Probing depth in monocular images. *Biological Cybernetics*, **56**, 355-366.

Stevens, K. A., & Brookes, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research*, **28**, 371-386.

Todd, J. T., & Reichel, F. D. (1989). Ordinal structure in the visual perception and cognition of smooth surfaces. *Psychological Review*, **96**, 643-657.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Wolfram, W. (1988). *Mathematica, a system for doing mathematics by computer*. Redwood City, CA: Addison-Wesley.