Perception & Psychophysics
2004, 66 (7), 1246-1259

Evidence for patchwork approximation of shape primitives

QUOC C. VUONG and FULVIO DOMINI
Brown University, Providence, Rhode Island

and

CORRADO CAUDEK
University of Trieste, Trieste, Italy

Investigators have proposed that qualitative shapes are the primitive information of spatial vision: They preserve an approximately one-to-one mapping between surfaces, images, and perception. Given their importance, we examined how the visual system recovers these primitives from sparse disparity fields that do not provide sufficient information for their recovery. We hypothesized that the visual system interpolates sparse disparities with planes, resulting in a patchwork approximation of the implicitly defined shapes. We presented observers with stereo displays simulating planar or smooth curved surfaces having different curvatures. The observers' task was to detect whether dots deviated from these surfaces or to discriminate planar from curved or planar from scrambled surfaces. Consistent with our hypothesis, increasing curvature had detrimental effects on observers' performance (Experiments 1–3). Importantly, this patchwork approximation leads to the recovery of the proposed shape primitives, since observers were more accurate at discriminating planar-from-curved than planar-from-scrambled surfaces with matched disparity range (Experiment 4).

The general goal of vision is to identify the perceptual and physiological mechanisms that allow human observers to successfully navigate through their environment, discriminate objects that populate it, and gather these objects for survival. Such actions are possible only if the appropriate information for carrying them out is somehow picked up from the environment by the visual system. What makes performing these actions difficult is that the same retinal images are consistent with the projection of many 3-D Euclidean scenes and objects. Given this problem, recent empirical and theoretical investigations have attempted to establish the appropriate spatial primitives for describing objects, images, and percepts. In particular, Lappin and Craft (2000) have addressed this problem by introducing a new formulation of visual information. They defined *visual information* as a set of spatial relationships specified intrinsically among texture elements belonging to discrete positions within a smooth 2-D manifold, without making reference to any extrinsic frame of reference. The central problem in surface perception, according to Lappin and Craft, is to identify the relationships that preserve one-to-one mappings between environmental objects, retinal images, and visual perception. Thus, they examined spatial relationships of different levels of complexity, depending on the number of texture elements involved[1] and the number of spatial dimensions along which these relationships are measured.

Special emphasis is given by Lappin and Craft (2000) to fourth-order structures—that is, to second-order spatial differentials taken in the 2-D neighborhood surrounding a given point (see also Koenderink & van Doorn, 1992). These higher order differential structures map one-to-one to local shapes, of which there are four alternatives: planar, parabolic, elliptic, or hyperbolic shapes (see Koenderink's, 1990, concept of *shape index*). The local shape at any point on a smooth surface depends only on the relative magnitude of two curvatures (i.e., second-order differentials) measured conjointly along two spatial dimensions in the local neighborhood surrounding the point, suggesting that local shape is the primitive information for spatial vision (see also Koenderink, 1990; Koenderink & van Doorn, 1992; Perotti, Todd, Lappin, & Phillips, 1998). That is, local shape is not derived from lower order spatial information (positions of texture elements, 2-D distances between texture elements, 3-D orientations, and so on).

Lappin and Craft (2000) tested their analysis by disrupting lower order structures while maintaining (higher) fourth-order structures. Here, we focus on their third experiment, which partially motivated our present work. In that experiment, Lappin and Craft measured hyperacuity on a task in which observers adjusted a probe dot to lie

This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) postgraduate scholarship to Q.C.V., and National Science Foundation Grant 78441 to F.D. We thank Bill Warren, Joe Lappin, Jim Todd, Allison Sekuler, and an anonymous reviewer for thoughtful comments and discussions on a previous version of this article. Correspondence should be addressed to Q. C. Vuong, Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany (e-mail: quoc.vuong@tuebingen.mpg.de).

coincident with planar or spherical surfaces defined by the spatial displacements of sparse texture elements, either between frames in the case of motion, or between the two monocular images in the case of stereovision. Since spatial relationships among the surface curvature in different directions are invariant across transformations such as expansion, translation, 2-D rotation, and depth rotation, which conversely disrupt zero-, first-, and second-order spatial relationships, Lappin and Craft hypothesized that visual acuity should be robust under such transformations. Their empirical results, obtained with both motion and disparity displays, support this conclusion. Indeed, observers were able to position the probe dot with a high degree of consistency (the standard deviation for individual observers across trials was less than 20 arc sec, which is in the hyperacuity range). Given the high degree of precision with which observers performed the adjustment task, the authors argued that observers were sensitive to fourth-order structures rather than deriving local shapes from more primitive structures such as spatial positions or depth gradients. That being said, Lappin and Craft leave open how smooth shape primitives are recovered, particularly since their stimuli do not provide a continuous disparity or motion field.

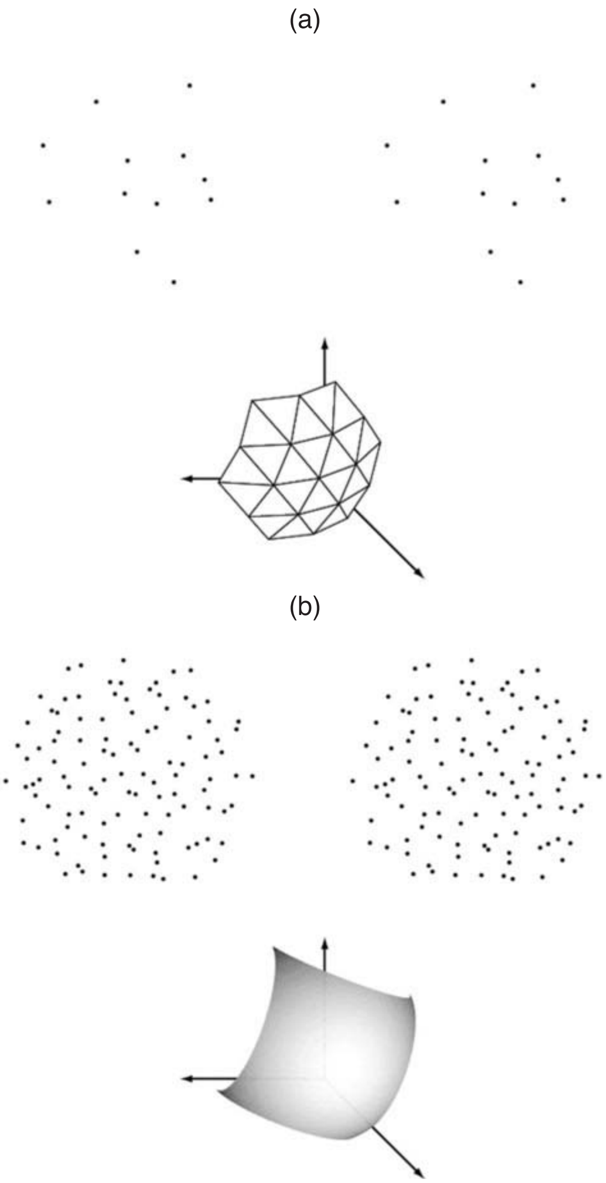

Figure 1. Two stereograms of the same simulated curved surface differing only in the number of texture elements. (a) With few texture elements, the percept is of a faceted surface approximating a curved surface as shown in the bottom panel. (b) With many more texture elements, the percept is of a smooth spherical surface, as shown in the bottom panel.

In the present article, we address the critical issue of how fourth-order differential structures can be estimated from a sparse disparity field that implicitly defines a smooth surface. This issue is important because, by definition, a continuous disparity field is necessary to implement Lappin and Craft's (2000) differential analysis. Therefore, some sort of interpolation process is needed to estimate a continuous disparity field from sparse data. Our primary goal was to establish whether the visual system interpolates sparse disparity fields with smooth surfaces. To illustrate this critical point and to further motivate our present experiments, consider Figure 1. The top panel of Figure 1a shows a stereogram of a simulated curved surface defined by a very small number of texture elements. The *same* simulated surface is shown in the top panel of Figure 1b, but with many more texture elements. In the first case, the percept is of a faceted surface, which *approximates* a smooth curved surface as depicted in the bottom panel of Figure 1a. On the other hand, with more texture elements, the percept is of a relatively smoother surface, as depicted in the bottom panel of Figure 1b. The important point is that when there are very few texture elements, many interpretations of the sparse display are possible other than the spherical surface that was actually simulated. For example, the sparse disparity field could be perceived as a distribution of points in space with no structure. That we still have a percept of a curved surface (albeit only approximately) with very few texture elements is in line with Lappin and Craft's proposal that local shapes are the primitive information of spatial vision (see also Koenderink, 1990; Koenderink & van Doorn, 1992; Perotti et al., 1998).

Following Lappin and Craft (2000), we measured observers' visual acuity at detecting whether probe dots deviated from otherwise smooth surfaces, which are implicitly specified by the sparse disparity field produced by a small number of discrete points sampled from those surfaces. To detect deviant probe dots with a high degree of visual acuity, we assume that observers estimate a continuous disparity field constituting the implicit surface in order to assign depth values to the probe dots, since these are positioned in regions with no disparity information. Depending on the precise nature of this in-

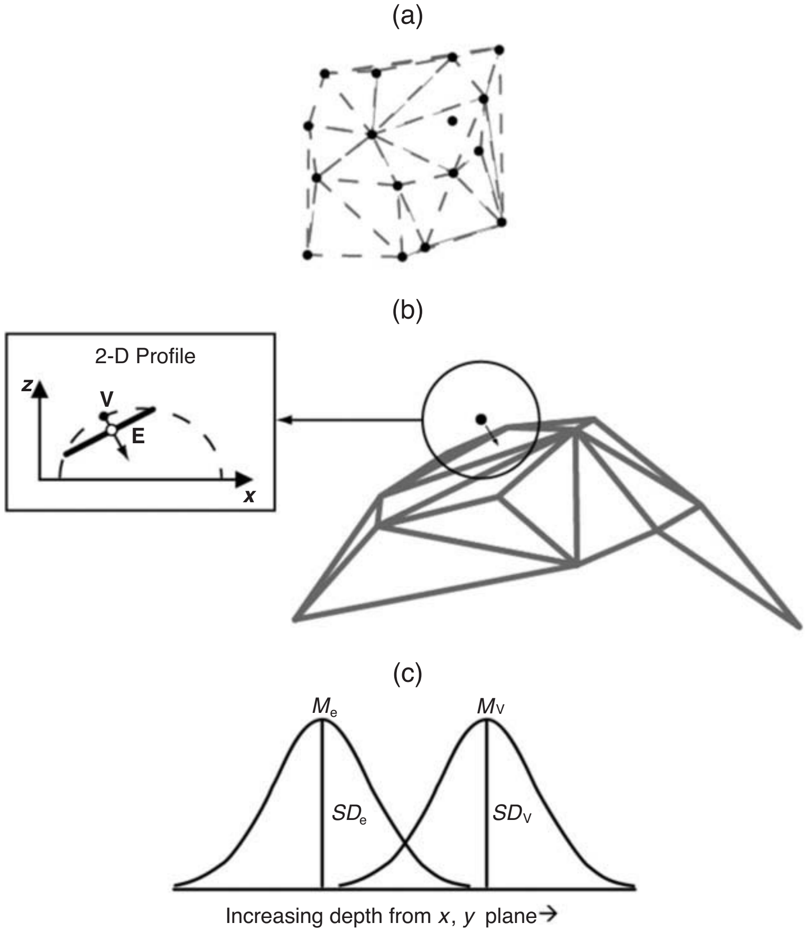

Figure 2. A model that implements the Turner et al. (1995) patchwork analysis of sparse data. The model takes 3-D coordinates as input. (a) It then performs a Delaunay triangulation of the input, excluding a probe point (the one that does not constitute a vertex of any triangle). (b) The probe point is projected onto the nearest plane derived from the triangulation. As shown by the inset, the estimated depth of the probe point from this projection (white circle marked E) is shallower than the veridical depth (filled black circle marked V), relative to the *x*, *y* plane. (c) Because the input varies from trial to trial, there is a distribution of veridical and estimated depth values with means M_v and M_e, and standard deviations SD_v and SD_e.

terpolation process, observers' acuity at detecting deviant probe dots may be significantly affected by properties of the disparity field independent of the implicit surface it specifies, such as the amount of curvature, the 2-D configuration of texture elements, or the number of texture elements.

For our study, we focused on the amount of curvature of the simulated surface. If the visual system interpolates a sparse disparity field with smooth surfaces (e.g., quadratic surfaces), this manipulation should not affect visual acuity since local shapes with different amounts of surface curvatures are, by definition, equivalent (Dijkstra, Snoeren, & Gielen, 1994; Koenderink, 1990; Koenderink & van Doorn, 1992; Lappin & Craft, 2000). Alternatively, Turner, Braunstein, and Andersen (1995) suggested that the visual system approximates implicitly smooth surfaces, which are affected by surface curvature and the number of texture elements. Their evidence suggests that the visual system fits planar patches to local texture elements to discriminate (simulated) smooth surfaces from noise in sparse structure-from-motion displays. As we show in the simulations below, if the visual system fits planes to a small number of texture elements to estimate an implicitly smooth surface, this patchwork approximation becomes less and less appropriate as the amount of surface curvature increases. Intuitively, if three sample points on the implicit surface have large depth differences, the plane that interpolates the points necessarily underestimates the (average) depth of the region. Clearly, the magnitude and variability of this underestimation depend on the amount of surface curvature and how the points are configured. Given the two percepts in

Figure 1, in the present work, we applied a similar patchwork approximation to the problem of recovering surfaces from sparse random-dot stereo displays.

Simulations

In this section, we show that the amount of surface curvature can affect the response of a model that implements the Turner et al. (1995) patchwork analysis of sparse data. As a secondary goal, we show that the configuration of points also affects the model's response. We do not intend for the simulations to justify any particular theory. Rather, we use the results to make qualitative predictions about the effect of surface curvatures on observers' performance on our tasks. Although the simulations involve only a single probe dot (Experiments 1 and 2), the model can be extended to multiple probe dots (Experiments 3 and 4).

The model is illustrated in Figure 2. The input to this model is a set of 3-D coordinates randomly sampled from a smooth curved surface. The model operates on this input as follows. It first performs a Delaunay triangulation on the 2-D coordinates of the input (Figure 2a). A probe dot is then randomly sampled on the implicit surface. The model's response consists of projecting this dot along its surface normal onto the nearest plane derived from the triangulation[2] (Figure 2b). Since the curved surface is approximated with a polygonal mesh of planes, the estimated depth of the probe dot is necessarily less than the veridical depth value when the dot is on the implicit surface. Moreover, given a small number of texture elements, this underestimation necessarily increases with the amount of surface curvature. By comparison, if a model interpolates sparse data points with smooth surfaces, the two depth values are necessarily equal for all surface curvatures.

For the simulations, we measured the model's response to smooth quadratic surfaces (Equation 1) with varying amounts of surface curvatures. For each such surface, 16 discrete points were positioned on the surface. The points were first placed in a regular grid pattern. The *x*- and *y*-coordinates of each point were then randomly jittered by an amount sampled from a uniform distribution with a mean equal to some proportion of the distance between adjacent grid points. One point was randomly selected as the probe point. Since both the probe and nonprobe points are randomly sampled on a trial-by-trial basis, the probe point has a distribution of veridical depth values with mean M_v and standard deviation SD_v, as well as a distribution of estimated depth values with mean M_e and standard deviation SD_e (Figure 2c).

We simulated 100 trials for each combination of surface curvature and amount of jitter and then computed the standard deviation in the model's *response error*, here defined as the probe point's veridical depth value minus its estimated depth value. An underestimation of depth value would result in positive response errors. The critical result of the simulations is shown in Figure 3a: Variability in the model's response systematically increases with the amount of surface curvature. As expected, when the sampled points were less jittered, the model's response was less variable. The model also underestimated the veridical depth value, and this underestimation increased with the amount of surface curvature (Figure 3b).

We conjectured that variability in the model's response could, in turn, lead to detriments in its ability to detect whether the probe point is on the implicit surface. To test this conjecture, we assume that the model's accuracy de-

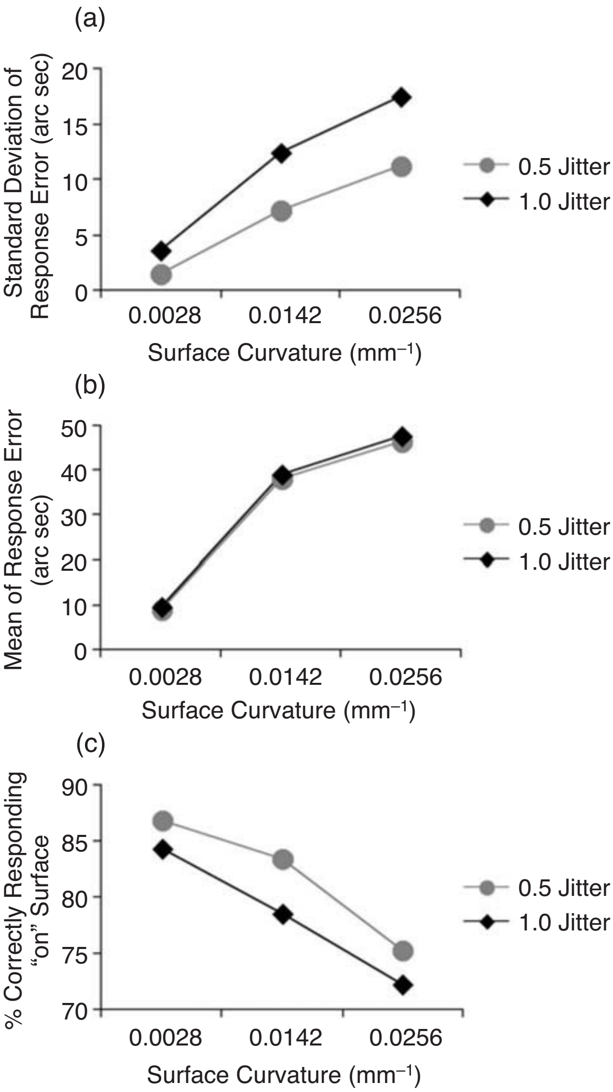

Figure 3. The results of the simulations. (a) The standard deviation of the model's response error (veridical depth − estimated depth) as a function of surface curvature, computed over 100 trials. The gray line indicates small 2-D jitter (a maximum value of half the distance between adjacent grid points), and the black line indicates large 2-D jitter (a maximum value of the distance between adjacent grid points). (b) The mean underestimation of the model's response as a function of surface curvature. (c) The response accuracy of the model to respond that the probe point is "on" the surface, as a function of surface curvature. The accuracy is based on the overlap of the distributions of the veridical and estimated depth values of the probe point.

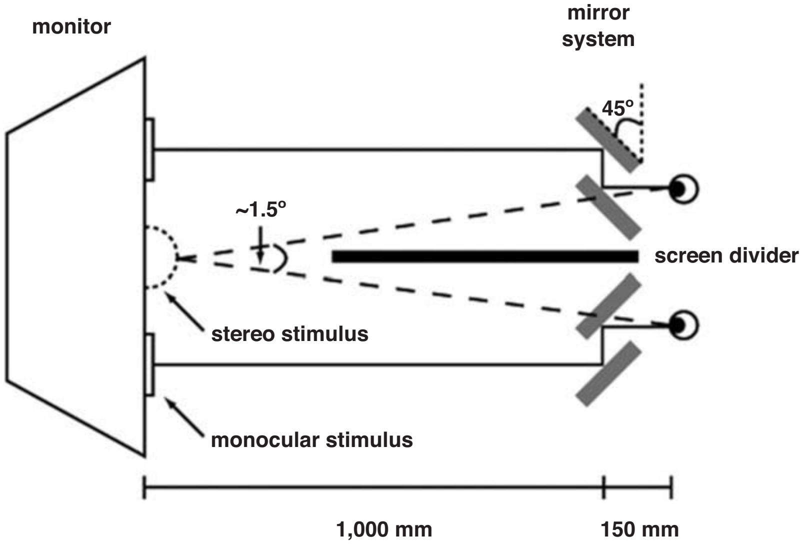

Figure 4. The experimental apparatus used in the present study. Through a system of mirrors, two monocular images are fused into a 3-D surface. Note that the figure does not show a screen that was used to divide the two monocular images and a chin-rest to stabilize an observer's head movements.

pends on the two distributions of the probe point's depth value. In effect, the more these two distributions overlap (Figure 2c), the more likely the model will correctly respond that the probe point is on the surface. The distributions derived from the simulations above were transformed into percentages that the model would correctly decide that the probe point is, in fact, on the surface. The uncertainty of this decision was modeled as Gaussian noise. The model's accuracy was then determined by means of the term [$1.0 - \mathrm{CDF}(M_v, M_e\ (SD_v + SD_e)/2)$], where CDF is the normal cumulative–distribution function. The critical result of this analysis is shown in Figure 3c: The model's accuracy systematically decreases as the amount of surface curvature increases.

To summarize, in four experiments, we tested whether the amount of surface curvature affected observers' ability to detect probe dots that deviated from implicitly smooth surfaces. Given the results of the simulations, if observers estimate a continuous disparity field from a sparse one via a patchwork-approximation process (Figure 2), their performance on this hyperacuity task would be negatively affected by the amount of surface curvature (Experiments 1 and 3). A secondary prediction, given the results of the simulations, is that observers would underestimate the depth of the probe dot (Experiment 2). In contrast, if observers interpolate a sparse disparity field with a smooth quadratic surface, for example, their performance should not be affected by the amount of surface curvature. Importantly, we provide evidence, consistent with Lappin and Craft's (2000) definition of visual information, that shape primitives can be recovered by a patchwork approximation of sparse disparity fields (Experiment 4).

EXPERIMENT 1

In Experiment 1, observers judged whether or not a probe dot was on a planar or curved surface. This task is a modified version of the adjustment task used by Lappin and Craft (2000). Here, however, both nonprobe *and* probe dots were randomly positioned on the image plane on each trial. The random sampling of positions on the implicit surface produces variability in the depth difference between the veridical depth on the implicit surface and the estimated depth, assuming a patchwork-approximation process (Turner et al., 1995). On the basis of the simulations above, we predicted that this variability would result in poorer performance on the acuity task for curved surfaces.

Method

Participants. The participants in Experiment 1 were 3 experienced psychophysical observers, including two of the authors (F.D. and Q.C.V.). The third experienced observer (D.R.) had participated in previous studies involving stereo displays, but was naive to the purposes of the present study.

Stimuli. The stimuli consisted of random-dot stereograms depicting quadratic surfaces oriented in depth. Spatial positions on these surfaces were sampled by a small number of high-luminance dots (100 cd/mm^2) on a low-luminance background (0.4 cd/mm^2). The curvature of the surfaces was defined by a parameter, C, with $C = 0.0\ \mathrm{mm}^{-1}$ for planar surfaces. The z value (i.e., simulated depth) of each dot was calculated from the following quadratic equation:

$$z = Cx^2 + Cy^2. \quad (1)$$

The general procedure used to generate these surfaces was as follows. Dots were first placed at the intersections of an $n \times n$ grid. The grid was circumscribed within a circle with a given radius (which eliminated the corner dots on the grid). The horizontal (x)

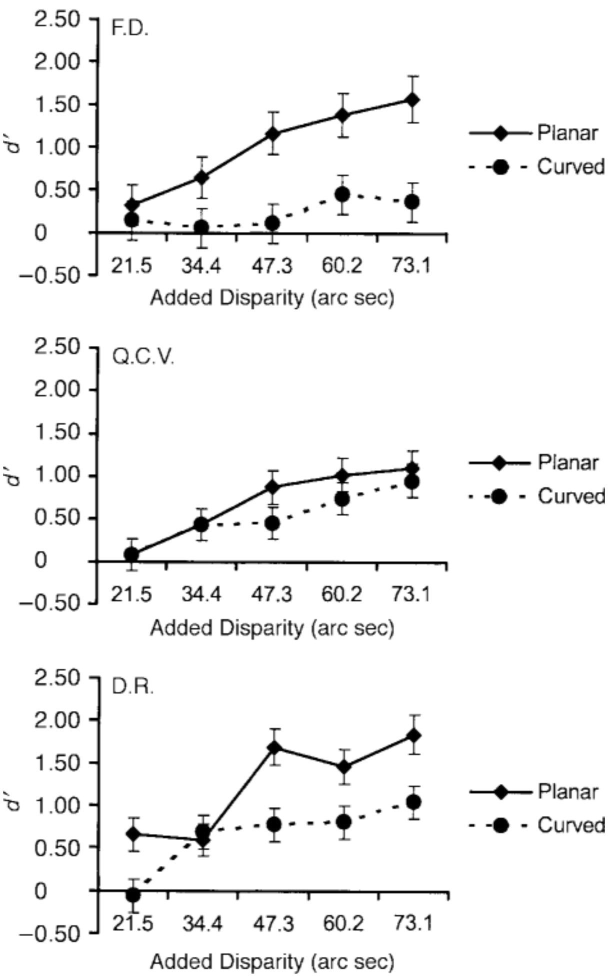

Figure 5. Individual d' scores obtained in Experiment 1 as a function of surface type (planar, curved) and added disparity (in arc sec). Error bars show individual standard deviation of d' scores (Marascuilo, 1970; Equation A1).

and vertical (y) positions of each dot were then perturbed in the image plane by a small random amount sampled from a uniform distribution (maximum 0.2º). This procedure distributed the dots evenly across the area of the circle. The depth map (z values) was calculated with Equation 1 above. The coordinates of the dots were then rotated first about the x-axis and then about the z-axis (line of sight). Lastly, we created a disparity map from the resultant depth map with the following equation:

$$\delta = \frac{2Iz}{D+z}, \tag{2}$$

where δ is the disparity value of the dot, z is the depth value calculated from Equation 1, I is the interocular distance, and D is the viewing distance. For all the experiments reported in this study, we fixed I = 60 mm (average interocular distance of the observers tested in the present study) and D = 1,150 mm. Two monocular images were generated from the disparity map and presented to observers using the apparatus described below.

The stereo displays used in Experiment 1 consisted of 12 dots, generated as described above. The radius of the circle (i.e., the maximum x and y values from the center of the stimulus) was 16.5 mm (0.8º) prior to any perturbation. In this experiment, simulated surfaces were either planar ($C = 0.0\ \mathrm{mm}^{-1}$) or curved ($C = 0.113\ \mathrm{mm}^{-1}$). This value produced a maximum simulated depth of 30.9 mm and a corresponding maximum horizontal disparity of 563.9 arc sec with respect to the screen when the surface was frontal parallel. For curved surfaces, the maximum simulated depth was at the perimeter of the circle. All surfaces were slanted 45º away from the frontal-parallel plane, and then randomly tilted from +45º to −45º about the line of sight. The stimuli remained on the screen until the observers responded.

The probe dot was randomly selected from one of the 12 dots initially generated. A briefly presented (250 msec) green circle indicated which dot served as the probe dot for that trial. The remaining 11 dots served as the nonprobe dots specifying the surface. To perturb the probe dot off the surface, we added a small amount of disparity to it by shifting its horizontal position in each monocular image.

Apparatus. Figure 4 provides a schematic illustration of the experimental apparatus used in all the experiments reported here. A high-resolution monitor (1, 600 × 1,280 addressable locations) was placed approximately 1,000 mm away from a mirror system. The observers sat approximately 150 mm away from the mirror system. Thus, the distance from the screen to the observer's eye was approximately 1,150 mm.

The two monocular images of each stimulus were displayed on the high-resolution monitor. A standard antialiasing procedure was used to achieve subpixel positional accuracy of each dot that comprised the monocular images (e.g., Domini & Caudek, 1999). Each dot consisted of a 5 × 5-pixel square encompassing the continuous-valued horizontal and vertical position calculated for that dot. Thus, each dot subtended approximately 0.03º × 0.03º. The (fractional) distance of each pixel from this "centroid" was used to scale its maximum luminance using a Gaussian weighting function with parameters μ = 0.0 and σ = 0.8. The left and right images were presented to the corresponding eyes of an observer using the mirror system. For each observer, the distance separating the two images on the screen was calibrated so that the vergence angle of the observer's eyes was approximately 1.5º.

As can be seen in Figure 4, the mirror system consisted of a set of four mirrors positioned so that each monocular image projected to the corresponding eyes. The mirrors were calibrated to be perpendicular to the ground plane (so that there were no vertical disparities), and to ensure that the rays from each monocular image were parallel throughout the path from the monitor to the eyes (see solid block lines in Figure 4). A screen was placed perpendicular to the monitor between the two monocular images to ensure that each eye did not see any portion of the other image (not shown in Figure 4). A chinrest was used to position each observer's head correctly and to restrict head movement (also not shown in Figure 4). For each observer, the height of the chinrest was adjusted so that his or her eyes were level with the vertical center of the mirrors.

The stimulus displays, response collection, and trial sequences were controlled by a PC computer. Responses were collected using a mouse. All experiments were conducted in a dark room.

Design and Procedure. There were two within-subjects factors: surface type (planar, curved) and added disparity (21.5, 34.4, 47.3, 60.2, 73.1 arc sec). These disparity values were added to the simulated disparity value of the probe dot when it was on the surface (i.e., veridical simulated depth), thereby perturbing it off the surface. The probe dot was perturbed either closer toward the observer or farther away from the observer than its veridical simulated depth, randomly determined on a trial-by-trial basis. The 10 surface type × added disparity conditions were blocked, and either 60 signal trials (F.D.) or 90 signal trials (Q.C.V. and D.R.) were collected in each condition (with an equal number of noise trials for each observer).

Prior to running the first session, all observers were briefly familiarized with the displays and the mouse buttons. On each trial, the observers were presented with a random-dot stereo stimulus with a green circle indicating the location of the probe dot (as described above). Their task was to judge whether the probe dot was on or off the surface defined by the remaining sample dots, and to press the appropriate mouse button (either the left or right button) to indicate their judgment. Their response cleared the screen, and the next trial followed 1,000 msec after stimulus offset. They could press the middle mouse button at any time during the stimulus presentation to briefly redisplay the circle. There was a self-timed break at the end of each experimental block.

All observers participated in two (F.D.) or three (Q.C.V. and D.R.) sessions, each lasting approximately 40 min to 1 h. There were 10 blocks in each session, with the order of blocks randomized across sessions. Each block consisted of 60 trials of either planar or curved surfaces. Half the curved surfaces were concave, and half were convex. At the beginning of each block, the observers were informed whether that block consisted of planar or curved surfaces. On half the trials, the probe dot rested on the surface (noise trials). On the other half, the probe dot was perturbed from the surface, always by the same added disparity for that block (signal trials).

Results and Discussion

Figure 5 plots the d' scores obtained for each observer in Experiment 1 as a function of added disparity, for planar and curved surfaces. For each observer, d' was computed from the proportion of hits (responding "off" when added disparity > 0.0 arc sec) and false alarms (responding "off" when added disparity $= 0.0$ arc sec) in each of the 10 conditions (2 surface types $\times$ 5 added disparities) pooled across the separate sessions. In this and all subsequent figures, error bars show the estimated standard deviations of the d' scores based on the formula given by Marascuilo (1970; see Appendix).

We compared the d' scores for the two surface types for all levels of added disparity across the 3 observers. There was a significant difference in the mean d' scores for planar and curved surfaces pooled across the 3 observers [$t(28) = 3.15$, $p = .004$; $M = 1.00$ and $SEM = 0.31$ for planar surfaces; and $M = 0.47$ and $SEM = 0.20$ for curved surfaces]. There were no systematic effects of concave or convex curved surfaces on the observers' d' scores. Since the probe dot was randomly selected on each trial, the observers' performance may also have depended on its position on a given trial. In Figure 6, we give the d' scores computed for each of the 12 possible probed positions for planar and curved surfaces. The d' scores were computed over all levels of added disparity across all 3 observers (for enough data points at each position). There does not appear to be any systematic effect of position; overall, the mean d' scores across observers are again larger for planar surfaces ($M = 0.57$ and $SEM = 0.14$) than for curved surfaces ($M = 0.36$ and $SEM = 0.19$).

The critical finding in Experiment 1 was that observers were more accurate at detecting dots perturbed from planar surfaces than they were at detecting dots perturbed from curved surfaces by the same amount. Thus, our results are consistent with the hypothesis that the visual system uses a set of planes to approximate a continuous disparity field from a sparse disparity field, rather than interpolating the sparse disparity field with smooth quadratic surfaces.

Planar surface ($M = 0.57$, $SEM = 0.14$)

	0.67	0.31	
0.64	0.47	0.73	0.03
0.54	0.96	0.81	0.62
	0.42	0.66	

Curved surface ($M = 0.36$, $SEM = 0.19$)

	0.65	−0.17	
0.50	0.67	0.55	0.14
0.69	0.54	−0.23	0.54
	0.29	0.52	

Figure 6. Mean d' scores for each of the 12 possible probe dot positions computed across the 3 observers for planar and curved surfaces. The scores are spatially arranged according to the 4 $\times$ 4 grid pattern that was used to generate the stimuli, with larger disparity values away from the center of the grid. The mean and standard errors of the mean are also shown.

We note, however, that observers' visual acuity in Experiment 1 is much worse than that reported by Lappin and Craft (2000). They found acuities in the 10–20 arc sec range, whereas observers in our study required approximately 40–70 arc sec of added disparity to achieve $d' = 1.00$. Several differences in the present experiment could partially account for this discrepancy: A detection task was used, both probe and nonprobe dots were randomly positioned in the image plane on each trial, fewer dots were used (12 vs. 19), and a larger curvature was used. However, we stress that our main comparison was between planar and curved surfaces under our stimulus conditions. In the following experiment, we found that if some of these differences were taken into account, observers in our study could achieve visual acuities similar to those reported by Lappin and Craft.

EXPERIMENT 2

The purpose of Experiment 2 was to replicate the results of Lappin and Craft (2000), given the considerably lower visual acuity reported in our Experiment 1. In the present experiment, we used a task identical to that of these investigators, modifying their stimuli only slightly. The observers in this experiment adjusted the probe dot until it was on the implicit spherical surface. The probe

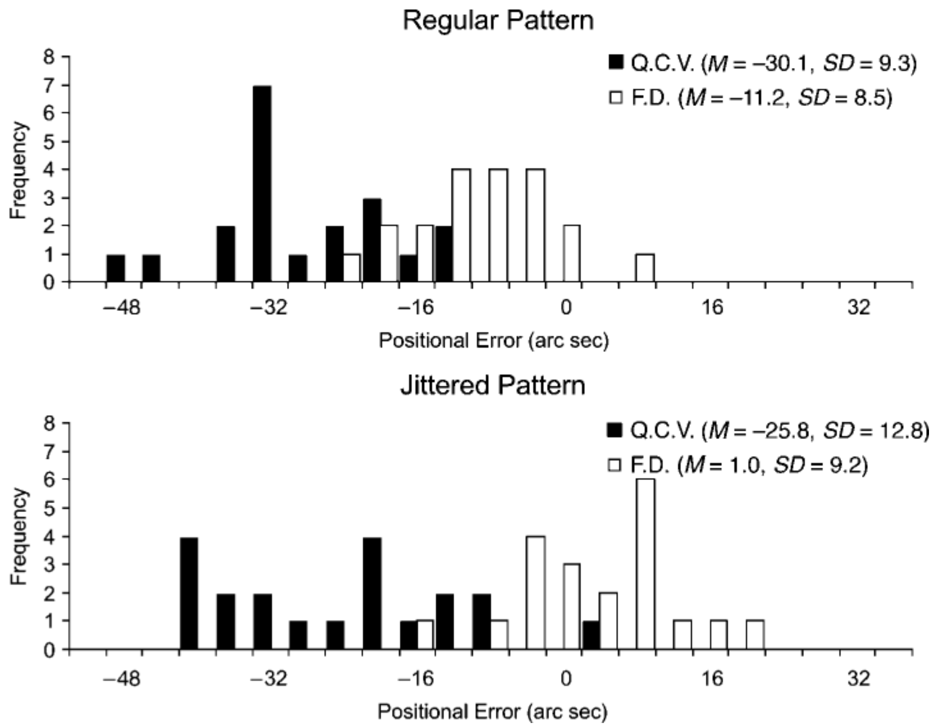

Figure 7. The frequencies of positional errors (in arc sec) in the probe adjustment task obtained in Experiment 2 for each observer and each stimulus pattern (regular, jittered). Also indicated are the means and standard deviations of these errors. Negative errors indicate that observers perceived the surface to be shallower than what is simulated.

dot was always positioned in the center of a stimulus pattern consisting of 18 dots arranged in a hexagonal array (as in Lappin & Craft, 2000) or in an irregular array (which varied from trial to trial). For this experiment, we also changed our dependent measure, given the nature of the task. Following Lappin and Craft, we measured observers' standard deviations at positioning a probe dot on the implicit surface. We also measured whether there was any systematic bias in the observers' adjustments. According to the simulations, the observers should underestimate the depth of the probe dot if the visual system uses some sort of patchwork approximation to estimate a continuous disparity field from sparse data.

Method

Participants. The first two authors (Q.C.V. and F.D.) served as participants in Experiment 2.

Stimuli. The 2-D stimulus patterns used in Experiment 2 were adapted from Lappin and Craft (2000). The first pattern consisted of 19 high-luminance dots on a low-luminance background arranged in a hexagonal pattern. The distance between any neighboring pairs of dots was approximately 0.75º. The maximum simulated depth value of the stimulus was 10.0 mm, with a corresponding maximum disparity value of 185.5 arc sec. The second pattern also consisted of 19 dots arranged in the same hexagonal pattern, but in which each dot was randomly jittered vertically and horizontally by a maximum amount of 0.3º. The only constraint was that the dot still fell within a radius of 1.8º after being perturbed in the image plane.

The stereo displays were generated as follows. The dots were orthographically projected onto an otherwise invisible spherical surface with a radius of 1.8º, which is slightly smaller than the stimuli used by Lappin and Craft (2000). In addition, the dots were randomly rotated about the line of sight from 0º to 50º, in 10º increments. The position of the surface was rotated 20º about the horizontal axis and then 20º about the vertical axis so that the surface normal pointed slightly down and slightly to the right. The probe dot was always at the center of the array.

Design and Procedure. The only factor manipulated in Experiment 2 was the stimulus pattern (regular vs. jittered). On each trial, the probe dot was displaced from the surface by a fixed amount of 20.9 arc sec (7.0 mm), plus an amount uniformly sampled from the range ±4.2 arc sec (±1.4 mm) along the surface normal. The observers moved this dot along the normal using the keyboard until they thought that the dot was coincident with the spherical surface, before advancing to the next trial with a mouse press. Both observers completed two blocks of 20 trials in one session lasting approximately 40 min (not including a break between blocks). Each block consisted of either regular-pattern or jittered-pattern stimuli. The order of the block was counterbalanced between the 2 observers. Observers took a long break (15–20 min) in between blocks to prevent adaptation.

Results and Discussion

Each observer provided 20 final probe positions (in arc sec) for the regular-pattern and the jittered-pattern stimuli. These final positions reflect deviations from the veridical position along the surface normal at the target position. Figure 7 plots the distribution of these positions for each observer and each stimulus pattern. The mean positional errors and standard deviations are also indicated in the figure.

The results show that the standard deviations (*SD*) of approximately 9–13 arc sec are within the hyperacuity range reported by Lappin and Craft (2000). Thus, when the probe dot was positioned in the center of the stimulus pattern, we were able to replicate Lappin and Craft's results. Clearly, the observers were able to perform the adjustment task with a high degree of stereoacuity with both stimulus patterns. However, our data tentatively suggest a minor improvement in performance when one computes the ratio of the *SD* of the regular-hexagonal pattern to the *SD* of the jittered pattern. Specifically, Q.C.V. was approximately 26% less variable with a regular-hexagonal pattern than with a jittered one. Likewise, F.D. was approximately 8% less variable with a regular-hexagonal pattern than with a jittered one. Indeed, the simulations above showed that the model responded more accurately with less jitter to the nonprobe points (compare the 0.5 jitter and the 1.0 jitter of Figure 3c). The smaller amount of 2-D jitter ultimately leads to less variability in the approximation across trials. This possibility is consistent with our recent finding that adding smooth monocular shading information to as few as nine sparse dots can drastically reduce the variability of observers' adjustments—relative to having only dots—by as much as 40% (Domini, Vuong, & Caudek, 2003). In this case, shading information may help constrain the perceived depth of the probe dot, thereby reducing the overall variability across trials.

Lastly, both observers underestimated surface curvature; that is, they perceived the depth of the probe dot to be shallower than its simulated depth value (for Q.C.V., $M = -28.0$ arc sec, and for F.D., $M = -5.1$ arc sec). This underestimation of the probe dot's depth follows from the model above: If observers position the probe dot to lie on a plane defined by any triplets of neighboring dots, the probe dot will necessarily have a shallower depth than that of the implicit curved surface. This underestimation is also consistent with the claim that the visual system cannot determine the veridical curvature of the projected surface (Perotti et al., 1998).

EXPERIMENT 3

In Experiment 1, we found that surface curvature affected observers' ability to detect whether a single probe dot was displaced from an otherwise smooth surface. However, only two surface curvatures were used. In Experiment 3, we systematically manipulated the amount of surface curvature to examine more closely the effect of curvature on surface perception. The observers' task in this experiment was to discriminate "smooth" curved surfaces from "noise" surfaces (e.g., Andersen, 1996; Norman, Lappin, & Zucker, 1991; Turner et al., 1995).

We presented our stereo stimuli very briefly in this experiment. Uttal, Davis, and Welke's (1994) work shows that observers can detect structure in stereo displays with presentations of less than 1 msec. Our brief stimulus presentation allowed us to test whether observers rely on global properties of the stereo display, such as the implicit surface, or on local cues, such as disparity discontinuities of neighboring texture elements.

Method

Participants. Three experienced psychophysical observers served as participants in Experiment 3. Again, Q.C.V. and F.D. served as observers. The 3rd observer (V.P.) was well practiced with the stereo displays used in the present experiment but was naive to its purposes.

Stimuli. The stereo displays used in Experiment 3 consisted of 32 dots. In this experiment, we simulated smooth surfaces with four different amounts of surface curvatures. All curved surfaces in this experiment were convex, and the maximum radius used in this experiment was 33 mm (1.6º).

Noise surfaces were generated as follows. First, a smooth surface was generated. Prior to rotating the generated surface in 3-D space, half of the 32 sample dots were randomly selected and displaced along their surface normal. The values of the selected dots, displacements were sampled from a normal distribution with $\mu = 0.0$ arc sec and three different σ values. The value of σ specifies the spread of the selected dots about the surface (in arc sec). We refer to this spread as the noise level. As in Experiment 1, all surfaces were first slanted 45º into the frontal-parallel plane, and then randomly tilted from +45º to −45º about the line of sight. In this experiment, stimuli were presented briefly for 250 msec.

In summary, in Experiment 3, there were three modifications to the detection task used in Experiment 1: (1) There were many more dots specifying the surfaces, (2) for noise surfaces, half the dots were randomly perturbed along their surface normal from a distribution of perturbations, and (3) the stimuli were presented very briefly.

Design and Procedure. There were two within-subjects factors: noise level (in standard deviations: 28.1, 35.1, and 42.1 arc sec for Q.C.V. and F.D.; and 70.3, 84.3, and 98.3 arc sec for V.P.) and surface curvature (0.0, 0.003, 0.009, and 0.014 mm^{-1}). When the simulated surface was frontal-parallel, these values simulated maximum depth values of 0.0, 3.1, 9.3, and 15.5 mm (relative to the screen), which corresponded to maximum disparity values of 0.0, 57.7, 172.3, and 285.7 arc sec.

The observers' task was to detect whether dots deviated from the otherwise smooth surfaces. All observers participated in four sessions, each lasting approximately 40 min to 1 h. There were four blocks in each session, and each block tested each surface curvature. The order of the four blocks was counterbalanced in a Latin square design to ensure that there were no practice or fatigue effects. Each block consisted of 120 trials. On half the trials, a smooth surface was presented to the observers (with a given magnitude of surface curvature). On the remaining trials, noise surfaces were presented to the observers, with an equal number of surfaces generated at one of the three different noise levels tested.

The 3 observers were briefly familiarized with the displays and the mouse buttons prior to starting the experiment. On each trial, the observers first saw a prestimulus display consisting of a square in each monocular image (with zero disparity). They fused the square to converge their eyes appropriately in preparation for the subsequently brief presentation of the stimulus proper (e.g., Uttal et al., 1994). After fusing the square, the observers initiated the stimulus presentation by pressing the middle mouse button. The stimulus was presented for 250 msec following the button press. The observers judged whether the briefly presented stimulus was a smooth surface (signal trial) or a noise surface (noise trial). They pressed the appropriate mouse button (either the left or the right button) to indicate their decision. The prestimulus square of the next trial followed 1,000 msec after their response. There was a self-timed break at the end of each experimental block.

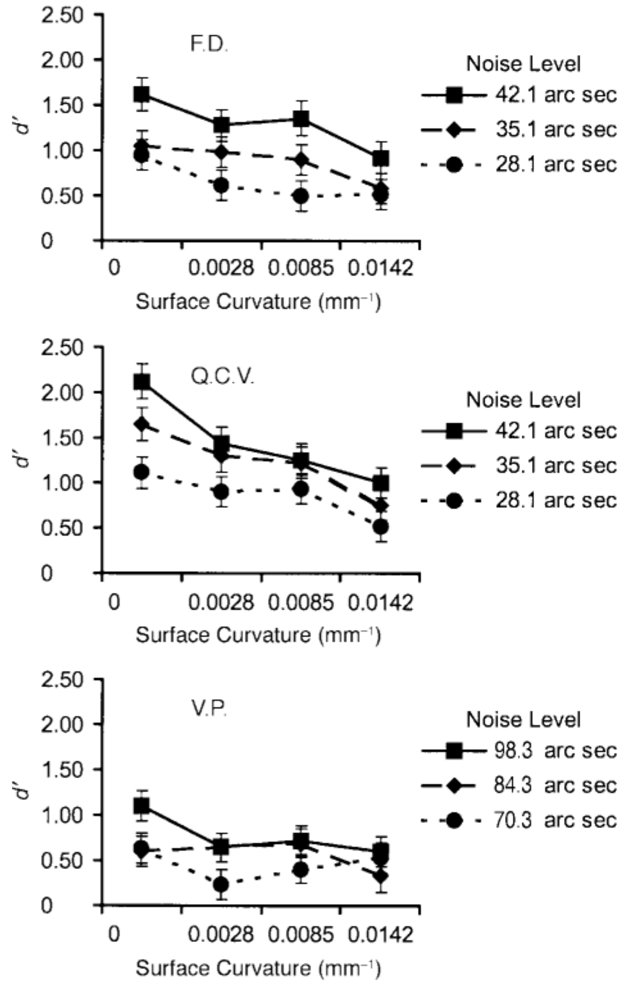

Figure 8. Individual *d′* scores obtained in Experiment 3 as a function of noise level and surface curvature. Error bars show individual standard deviation of *d′* scores (Marascuilo, 1970; Equation A1).

Results and Discussion

Figure 8 plots the observers' *d′* scores for each of the 12 conditions (3 noise levels $\times$ 4 surface curvatures) in Experiment 3. We presented the data in this manner for ease of comparison with the model's predictions, shown in Figure 3c, in which performance decreases as surface curvature increases. The *d′* scores were computed as follows. First, responses for the different curvatures were pooled across the four sessions. For each value, hits were computed over all signal trials ($n = 240$), and this value was used for all *d′* computations. False alarms were computed over all noise trials for each noise level separately ($n = 80$).

A repeated measures analysis of variance (ANOVA) with noise level and surface curvature as within-subjects factors was conducted on the *d′* scores. Both factors were significant: noise level [$F(3,6) = 19.54$, $p < .001$] and surface curvature [$F(2,4) = 9.95$, $p < .01$], as was their interaction [$F(6,12) = 3.14$, $p = .04$]. The results of Experiment 3 show that surface curvature systematically affected observers' accuracy at detecting displaced dots from briefly presented surfaces. At all noise levels, observers' ability to discriminate surfaces from noise stimuli decreased as curvature increased.[3] We believe that our results are consistent with those obtained in Experiment 1 and support the hypothesis that a patchwork-approximation process is used to recover local shape primitives from sparse displays (Turner et al., 1995). Given the brief duration of our stimuli (relative to that of those in Experiment 1), it seems unlikely that observers performed the task by the necessarily more time-demanding process of comparing disparity differences between neighboring dots. Indeed, the results of Experiment 3 suggest that observers are sensitive to a global property of the sparse disparity display that may, in turn, depend on the simulated local shape primitive.

EXPERIMENT 4

The results of Experiment 3, in which the stimuli were presented very briefly, suggest that observers may be sensitive to global properties of a sparse disparity field, such as the implicit surface specified by the sample dots. In Experiment 4, we further examined this sensitivity to possible global properties of the sparse disparity field. Unlike in Experiment 3, however, we used a range of surface curvatures that was near threshold detection, and we generated noise stimuli that provided a stronger and more direct test of this sensitivity.

In Experiment 4, observers were asked to discriminate planar surfaces from either smooth curved surfaces or from matched "scrambled" curved surfaces in which the disparities between visible texture elements were randomly permuted. The two types of stimuli used in this experiment are illustrated in Figure 9. In the top panels of Figures 9a and 9b, we show a 2-D cross section of a simulated curved surface (gray dashed line). In Figure 9a, the otherwise invisible surface is sampled by dots (solid black circles). In contrast, in Figure 9b, the depth values of the dots in Figure 9a are permuted to produce a scrambled version of the same surface. The bottom panel of Figures 9a and 9b show stereograms of these two types of surfaces.

As can be seen in Figure 9, the critical difference between these two stimuli is that the disparity field changes gradually across neighboring image regions for smooth surfaces, but abruptly for scrambled surfaces. However, both stimuli have the same disparity range, shown by the gray region in the figure. This disparity range, in turn, depended on surface curvature. Therefore, if observers are sensitive to some global property of the stimuli, we hypothesize that they would be more accurate at discriminating planar from smooth surfaces than planar from scrambled surfaces.

Method

Participants. The same 3 observers from Experiment 3 participated in Experiment 4 (Q.C.V., F.D., and V.P.). Again, V.P. was naive to the purposes of the present experiment.

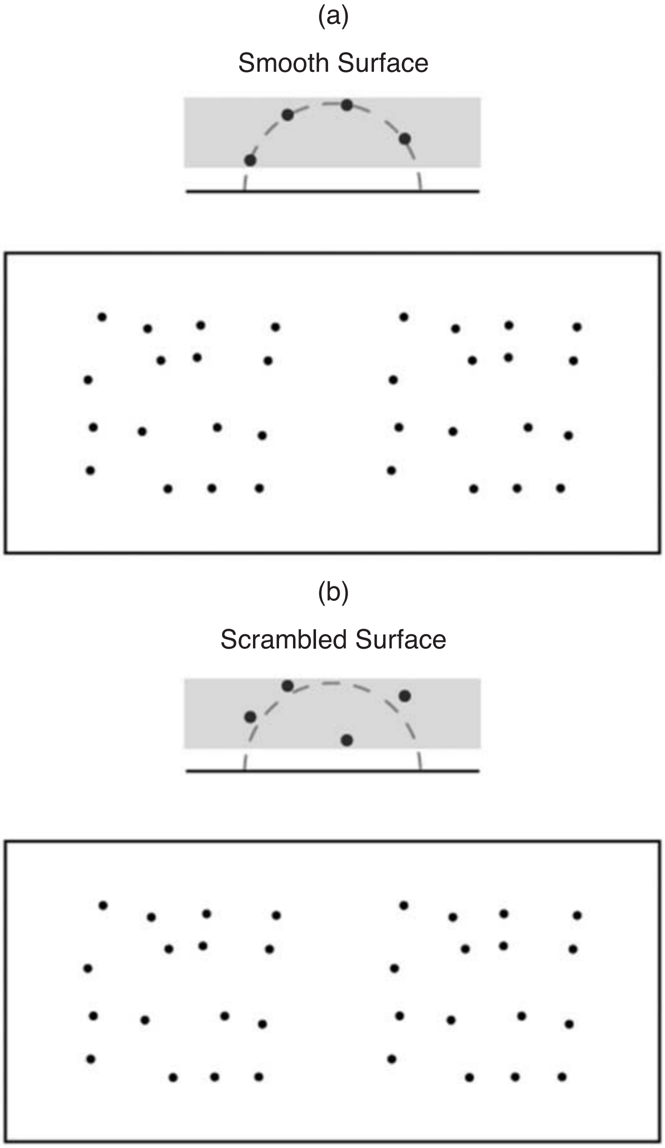

Figure 9. The two types of surfaces used in Experiment 4: (a) a smooth curved surface, or (b) a scrambled surface (derived from a smooth surface). The top panel of (a) shows a 2-D cross section of a smooth curved surface (dashed gray line) sampled by a sparse number of dots (black circle). The gray box represents the simulated depth range of the dots. Similarly, the top panel of (b) shows a 2-D cross section of a scrambled surface, in which the depth values of the sample dots (black circle) are permuted so that they no longer lie coincident with the underlying smooth surfaces (dashed gray line). However, as shown by the gray box, the depth range remains the same as in (a). In addition, the dots have the same horizontal position as in (a). The bottom panels of both (a) and (b) show stereograms of the two types of surfaces. Note the identical (*x*,*y*) position of each dot (except for slight shifts due to the disparity calculations of Equation 2).

Stimulus. The stereo displays used in Experiment 4 consisted of 16 dots. As in Experiment 3, these dots were circumscribed in the image plane within a circle with a radius of 33 mm (1.6º). In this experiment, the smooth curved surfaces had four near-threshold surface curvatures. The curvatures were determined individually for each observer in a pilot experiment, using similar stimuli and tasks. As in Experiment 1, there were an equal number of convex and concave surfaces presented to the observers.

The scrambled surfaces were generated as follows (see Figure 9). First, a smooth curved surface was generated. Before the surface

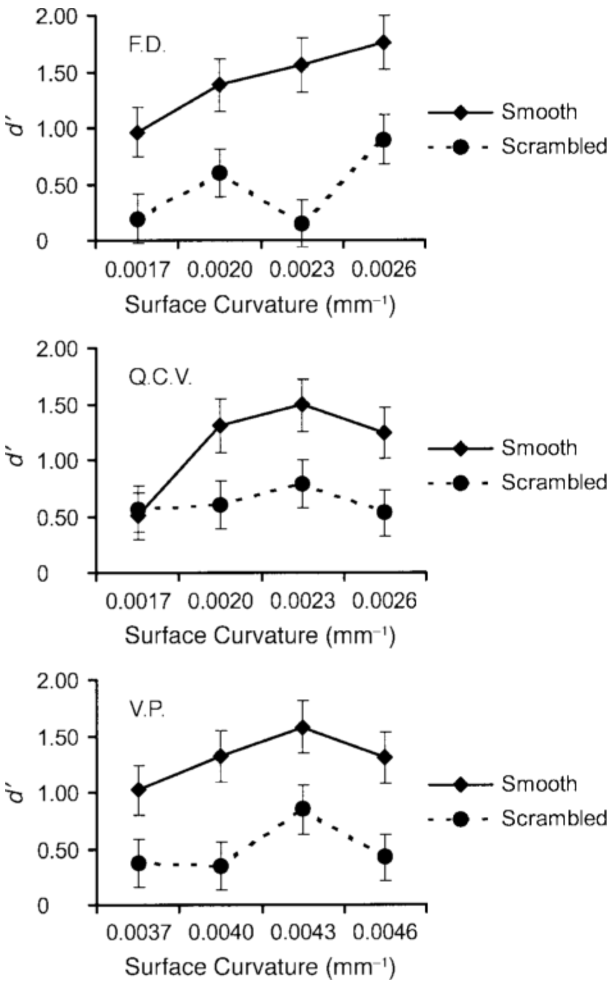

Figure 10. Individual d' scores obtained in Experiment 4 as a function of surface type (smooth, scrambled) and surface curvature (around threshold). Error bars show individual standard deviation of d' scores (Marascuilo, 1970; Equation A1).

was rotated in 3-D space, the computed disparity values of all the dots were randomly permuted between the dots with no constraints, so that the disparity field corresponded to a 3-D volume rather than a smoothly curved surface (e.g., Turner et al., 1995). This scrambling procedure preserves the distribution of disparity but introduces discontinuities between neighboring dots. Consequently, the mean difference between *all* pairwise disparity values is the same for both smooth and scrambled surfaces. For both types of surface, this mean difference ranged from 11.0 arc sec to 16.4 arc sec for Observers Q.C.V. and F.D., and from 23.7 arc sec to 29.0 arc sec for Observer V.P. However, local regions (i.e., neighboring dots) of scrambled surfaces necessarily have, on average, larger disparity differences than do local regions of smooth surfaces. The overall effect of the scrambling procedure is to produce a jagged or depth-jittered surface. As in the previous experiments, all surfaces were first slanted 45º in the image plane, and then randomly tilted from +45º to −45º about the line of sight. The stimuli were presented briefly for 250 msec.

Design and Procedure. There were two within-subjects factors: surface type (smooth, scrambled) and surface curvature (0.0017, 0.0020, 0.0023, and 0.0026 mm^{-1} for Q.C.V. and F.D., and 0.0037, 0.0040, 0.0043, and 0.0046 mm^{-1} for V.P.). These curvatures produced maximum simulated depths (as defined in Experiments 1 and 3) of 1.86, 2.17, 2.48, and 2.79 mm for Q.C.V. and F.D., and 4.02, 4.33, 4.64, and 4.95 mm for V.P. These simulated depth values, in turn, corresponded to maximum disparity values of 34.6, 40.4, 46.2, and 52.0 arc sec for F.D. and Q.C.V., and 75.0, 80.8, 86.6, and 92.2 arc sec for V.P. The eight surface type $\times$ surface curvature conditions were blocked, with 72 trials collected in each condition.

The observers were briefly familiarized with the displays and the mouse buttons prior to starting the experiment and were instructed to judge whether a briefly presented stimulus was a planar surface. They pressed the appropriate mouse button (either the left or the right button) to indicate their judgments. As in Experiment 3, at the beginning of each trial, the observers first saw a prestimulus square. After fusing the square, they initiated the stimulus presentation by pressing the middle mouse button. The stimulus was presented for 250 msec following the button press. The prestimulus square of the next trial followed 1,000 msec after their response. There was a self-timed break at the end of each experimental block.

All observers participated in two sessions, each with four blocks. Each session lasted approximately 40 min to 1 h. Across the two sessions, the eight experimental conditions were tested. The order of the eight blocks was randomly determined for each observer. Each block consisted of 144 trials. On half the trials, planar surfaces were presented to the observers (signal trial). On the remaining trials, nonplanar surfaces were presented to the observers (noise trial).

Results and Discussion

Figure 10 plots individual observers' d' scores for each of the eight conditions (2 surface types $\times$ 4 surface curvatures) in Experiment 4. The d' scores were computed from the hits (correctly judging a planar stimulus as planar) and false alarms (incorrectly judging a nonplanar stimulus as planar) obtained in each condition (72 signal trials and 72 noise trials).

An ANOVA with surface type and surface curvature as within-subjects factors was conducted on the d' scores. There was a significant main effect only of surface type [$F(1,2) = 34.30, p = .02$]. The critical finding in Experiment 4 was that observers were better at discriminating planar-from-curved surfaces than at discriminating planar-from-scrambled surfaces. Across the 3 observers, the mean d' score and standard errors of the mean for smooth surfaces were $M = 1.29$ and $SEM = 0.19$ and for scrambled surfaces, they were $M = 0.53$ and $SEM = 0.14$. This finding suggests that, in interpolating a sparse disparity field, the visual system does not restrict its analysis to local regions of the disparity field. Such an approach, in fact, would have favored the discrimination between planar and scrambled surfaces in which the disparity difference of neighboring dots was larger, on average, for scrambled surfaces than for smooth curved surfaces.

GENERAL DISCUSSION

The present research extends a previous investigation by Lappin and Craft (2000). For Lappin and Craft, the problem of spatial vision is that of identifying the spatial correspondences between environmental objects, their retinal projections, and their perceptions. They pointed out that an isomorphism exists between fourth-order differential structures of environmental surfaces and of the disparity or displacement fields of the retinal images, their important point being that, within this mapping, the information about local shape is fully preserved. That is, local shape is the spatial primitive of surface perception (see also, e.g., Perotti et al., 1998). Lappin and Craft presented strong evidence that the visual system is sensitive to the fourth-order differential structures of images, which, according to these investigators, map to percepts of smooth shapes.

However, given sparse disparity fields that randomly sample positions on simulated surfaces, our percepts do not correspond to smooth surfaces by default (see Figure 1). Indeed, Lappin and Craft (2000) did not explicitly address how shape primitives are recovered from sparse disparity or motion fields. This open issue was examined in the present study. On the basis of the work by Turner et al. (1995), we hypothesized that a patchwork-approximation process is used to estimate a continuous disparity field from sparse data to implement the Lappin and Craft analysis. This process is sensitive to, among other manipulations, surface curvature and the configuration of discrete texture elements. We believe that this patchwork-approximation process provides the simplest account that is consistent with the findings reported in the present study.

First, we found that the amount of surface curvature affected observers' visual acuity on tasks and stimuli similar to those used by Lappin and Craft (2000). In our study, observers were presented with stereo displays simulating planar or smooth curved surfaces having different curvatures. The general task was to detect whether dots deviated from these otherwise smooth surfaces. In Experiment 1, we found that observers were more accurate at detecting dots perturbed from planar surfaces than dots perturbed from curved surfaces by the same amount. This result was obtained with stimulus displays similar to those used by Lappin and Craft, but with the probe and nonprobe dots randomly distributed in the image plane. In Experiment 3, we extended the results of the first experiment by parametrically varying the amount of curvature of the simulated surfaces. We also presented the stimuli very briefly to reduce the likelihood that observers were comparing local disparities.

Second, in Experiment 2, we replicated Lappin and Craft's (2000) results with our stimulus apparatus. By forcing the probe dot to be always in the center of the stimulus configuration, we obtained hyperacuities in the range reported by those authors. Despite the high degree of visual acuity shown by the observers, we also found a slight improvement in performance when the dots were placed in a regular-hexagonal pattern as compared with a jittered one (see also Domini et al., 2003). We also found that observers underestimated the depth of the probe dot with respect to its veridical simulated depth. The results of Experiments 1–3 therefore conformed to the qualitative predictions of a model that implements the Turner et al. (1995) patchwork analysis of sparse data.

Perhaps the most intriguing results in the present study are those reported in Experiment 4. In that experiment, observers were more accurate at discriminating very briefly presented planar surfaces from curved surfaces than they were at discriminating planar surfaces from matched scrambled surfaces, even though a local analysis would have predicted the opposite results; that is, it would have favored the discrimination between planar and scrambled surfaces. These results suggest that the visual system may be sensitive to the simulated surface that was used to generate the disparity field, since the smooth and scrambled surfaces have the same range of local disparities. Indeed, this sensitivity is consistent with Lappin and Craft's (2000) claim that the visual system is very sensitive to higher order spatial relationships. This sensitivity to the simulated surface is also consistent with the recent results reported by Glennerster, McKee, and Birch (2002). They found that observers' thresholds for discriminating local disparities depended on reference to a simulated surface—in their case, a plane oriented toward or away from the observers. Our contribution, in this regard, is to provide evidence that a patchwork approximation can, in fact, lead to the recovery of shape primitives that preserve the (approximately) one-to-one mapping between objects in the environments, the retinal images, and their percepts, even for sparse visual information.

REFERENCES

Andersen, G. J. (1996). Detection of smooth three-dimensional surfaces from optic flow. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 945-957.

Dijkstra, T. M. H., Snoeren, P. R., & Gielen, C. C. A. M. (1994). Extraction of three-dimensional shape from optic flow: A geometric approach. *Journal of the Optical Society of America A*, **11**, 2184-2196.

Domini, F., & Caudek, C. (1999). Perceiving surface slant from deformation of optic flow. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 426-444.

Domini, F., Vuong, Q. C., & Caudek, C. (2003, September). *Evidence of strong coupling between shading and stereo information to 3-D surface reconstruction*. Poster presented at the 26th European Conference on Visual Perception, Paris.

Glennerster, A., McKee, S. P., & Birch, M. D. (2002). Evidence for surface-based processing of disparity. *Current Biology*, **12**, 825-828.

Koenderink, J. J. (1990). *Solid shape*. Cambridge, MA: MIT Press.

Koenderink, J. J., & van Doorn, A. J. (1992). Generic neighborhood operators. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **14**, 597-605.

Lappin, J. S., & Craft, W. D. (2000). Foundations of spatial vision: From retinal images to perceived shapes. *Psychological Review*, **107**, 6-38.

Marascuilo, L. A. (1970). Extensions of the significance test for one-parameter signal detection hypothesis. *Psychometrika*, **35**, 237-243.

Norman, J. F., Lappin, J. S., & Zucker, S. W. (1991). The discriminability of smooth stereoscopic surfaces. *Perception*, **20**, 789-807.

Perotti, V. J., Todd, J. T., Lappin, J. S., & Phillips, F. (1998). The perception of surface curvature from optical motion. *Perception & Psychophysics*, **60**, 377-388.

Turner, J., Braunstein, M. L., & Andersen, G. J. (1995). Detection of surfaces in structure from motion. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 809-821.

Uttal, W. R., Davis, N. S., & Welke, C. (1994). Stereoscopic perception with brief exposures. *Perception & Psychophysics*, **56**, 599-604.

NOTES

1. The number of texture elements determines the differential order of the spatial relationship: One element gives a zero-order relationship (position), two elements give a first-order relationship (length, orientation), and three elements give a second-order relationship (curvature). Higher orders are possible, but psychophysical data indicate that the human visual system only uses up to second-order information.

2. Considering only the x- and y-coordinates of the input, the *nearest triangle* is defined as the triangle whose vertices surround the probe point.

3. We acknowledge that interpretation of the results of Experiment 3 is limited by the design we used, as pointed out by a reviewer. In particular, we intermixed different noise levels within each block. Since observers are necessarily more confident about their responses with larger noise values and less so with smaller noise values, their response criterion may have differed across noise levels, thereby contaminating signal trials (in which no noise is added). We reiterate, however, that our main purpose was to compare performance across different surface curvatures, which was fixed within each block.

APPENDIX

The estimated variance of d' scores was computed as follows (Marascuilo, 1970):

$$SE^2 = \frac{p_1(1-p_1)}{n_1\left[f(Z_1)\right]^2} + \frac{p_2(1-p_2)}{n_2\left[f(Z_2)\right]^2}, \tag{A1}$$

where

p_1 = proportion of hits (or correct rejections),
p_2 = proportion of false alarms (or misses),
n_1 = number of signal trials,
n_2 = number of noise trials,

$$Z_1 = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{Z_i} e^{\frac{-1}{2x^2}}\, dx,$$

and

$$f(Z_i) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2Z_i^2}}.$$

(Manuscript received February 21, 2003;
revision accepted for publication December 8, 2003.)