Perceptual Localization of Surface Position

Flip Phillips and James T. Todd
The Ohio State University

Jan J. Koenderink and A. M. L. Kappers
University of Utrecht

In 2 experiments, observers were required to identify corresponding points on an object viewed from multiple orientations. On each trial, a surface was presented initially with a single target location marked by a small dot. Following a brief blank interval, the same surface was presented again at a different orientation. The observer was required to position an adjustable probe dot in this 2nd display to match the location of the target in the 1st view. Under optimal conditions, the variance in their settings over multiple trials was just a few minutes of arc, though these errors varied significantly with the structural complexity of the depicted surface.

Consider the following scenario. A young boy comes in from playing to have dinner with his family. His father at one end of the table looks over and says, "Jason, you have a smudge on your nose. Please wipe it off." His mother at the other end of the table nods approvingly, and then responds, "Yes dear, and while you are at it, you should wipe off the one on your chin as well."

Although it is common in day-to-day experiences for friends and loved ones to inform one another of blemishes that appear on different parts of their clothing or anatomy, from the standpoint of perceptual theory, this phenomenon is quite remarkable. In the scenario described above, for example, the parents are able to distinguish two smudges that have no other identifying characteristics except for where they are located on the child's face. These identities appear to be preserved, moreover, even though they are observed from different vantage points at opposite ends of the dinner table.

It is important to keep in mind when considering this phenomenon that the ability to localize positions on a surface from multiple vantage points is not unique to faces. Consider, for example, a surface cross section viewed from two different orientations as depicted schematically in Figure 1. In the upper part of this Figure, there is a small dot located in the valley between two bumps, but in the lower part it has been moved to a different location. Clearly in this case an observer would have no difficulty identifying that the dots are located in different positions, even though those positions are viewed from different orientations.

Note, however, that there are some objects encountered in

nature, for which all surface points are perceptually indistinguishable. In order to localize individual positions, a surface must have an appropriate level of structural complexity, at an appropriate scale to be detected by the human visual system (see Figure 2). The sphere and the plane are particularly extreme examples. Because they possess infinite symmetry, no one point could be distinguished from any other, unless there were some pattern of surface reflectance to provide a set of potential landmarks. Similarly, the ability to localize points may be impaired if the structure of a surface is too complex. A good example is a field of grass, which has an intricately complex structure, but at a scale that is too small to identify specific features under normal viewing conditions.

For surfaces with appropriate levels of structural complexity, there is strong anecdotal evidence to suggest that individual surface points viewed from different orientations can be perceptually localized with at least some degree of precision. This viewpoint invariance would be difficult to accomplish if smoothly curved surfaces were perceptually represented by depth or orientation maps, as is typically assumed in many theoretical analyses of 3-D form perception (e.g., see Marr, 1982). The problem with these representations is that depths and orientations are extrinsic surface properties defined relative to the point of observation, and they do not remain invariant when the point of observation is moved. In principle, the relative spatial positions of two points could be compared by mentally transforming one view so that it is aligned with the other (e.g., see Shepard & Cooper, 1986; Shepard & Metzler, 1971; Tarr, 1995), but this would require a six-dimensional search in the most general case to find the appropriate transformation parameters (i.e., 3 degrees of freedom for translation and 3 degrees of freedom for rotation).
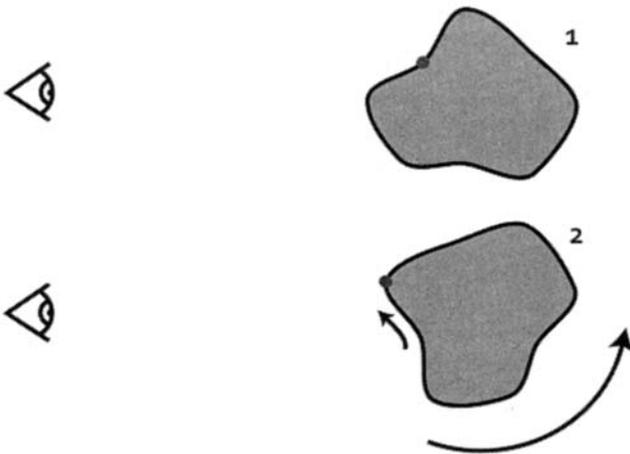
An alternative hypothesis, which is perhaps more plausible, is that the viewpoint-invariant identity of a given surface position must be defined by some intrinsic aspect of local surface structure that does not depend on an external frame of reference. For example, one possibility is that each local region of a surface could be perceptually represented by its two principal curvatures (see Koenderink, 1989), which are defined independently of any particular point of observation. Thus, if a surface moves relative to the

Figure 1. In Position 1 at the top of this illustration, a marker is placed in a valley between two "hills" on the object. In Position 2, the object has been rotated and the point has been moved. An observer would have little difficulty in detecting this change.

observer, or vice versa, the depths and orientations of each local region would change, but the magnitudes of their principal curvatures would remain invariant.

Although the localization of relative spatial position may be a fundamental process of human perception, there has been surprisingly little research on this topic. One notable exception is the work of DeValois, Lakshminarayanan, Nygaard, Schlussel, and Sladky (1990), who investigated the discrimination of relative spatial position on simple two-dimensional configurations of points and lines. For example, in one of their experiments, the displays all contained a pair of horizontal reference lines—a standard and a test—each of which had a small vertical bar positioned at slightly different locations. Observers were required to indicate whether the vertical bar in the test configuration was shifted to the right or left of the corresponding position of the vertical bar on the standard. Performance on this task and on other related ones was remarkably accurate. Weber fractions in all of the different conditions were approxi-
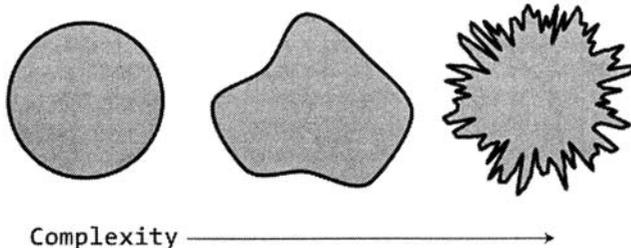
Figure 2. It would be impossible to remember a particular location on the circle to the left because all of the points look the same. The same problem applies to the jagged object on the right—each location looks pretty much like the others. The medium-complexity object in the middle lends itself most easily to the identification of various locations.

mately 0.02, even when the test configuration was rotated, reflected, or magnified relative to the standard.

The research described in the present article was designed to determine if similarly high levels of performance can be achieved at discriminating the relative spatial locations of points on smoothly curved surfaces viewed from different orientations. The paradigm used in these studies was conceptually quite simple. On each trial an observer was presented with a computer-generated surface patch defined by shading, texture, and binocular disparity, with a small probe dot to indicate a particular target location. Following a short blank interval, the same surface was presented again from a different viewing perspective, and with a different random texture so that the target could not be localized on the basis of the pattern of surface reflectance. The observer's task was to manipulate an adjustable probe to match the position of the target in the first view. The goals of this research were threefold: first, to measure the precision with which observers are able to localize the positions of individual surface points under a variety of conditions; second, to examine how this precision is influenced by the complexity of surface structure; and third, to measure how localization errors covary with different local properties of surface structure, such as depth, orientation, or curvature.
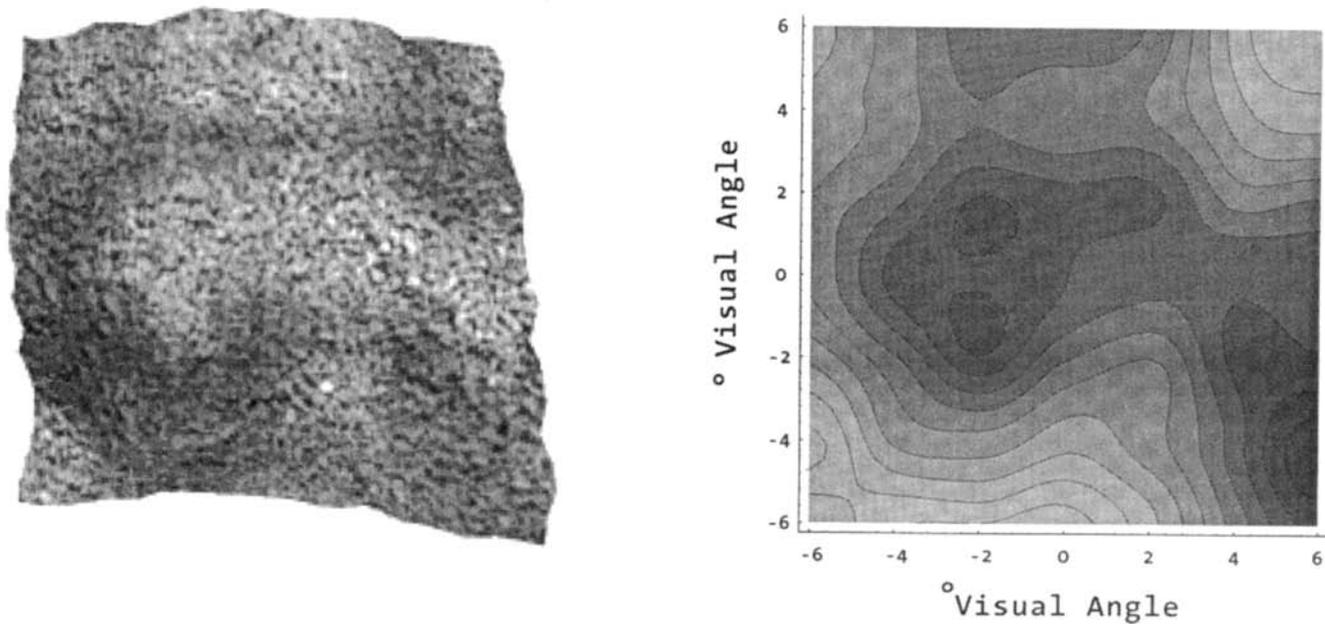
Experiment 1

Method

Apparatus. All stimuli were computer generated surfaces, displayed on a Silicon Graphics Crimson workstation. In all presentations, the surfaces were shown with texture and shading and were viewed stereoscopically using Crystal Eyes liquid crystal shutter glasses. Viewing distance was a constant 57.4 cm and the entire display screen subtended an angle of 34° horizontally and 28° vertically. The experiments took place in a semidarkened room with a chin rest used to maintain a constant viewing distance. Responses were provided using the mouse buttons.

Stimuli. The construction of the stimuli in this experiment is similar to that used by Phillips and Todd (1996). A turbulent surface was generated by combining multiple octaves of two-dimensional "noise" (as per Peachey, 1985, and Perlin, 1985) using a $1/f^\beta$ power spectrum. This type of fractal turbulence is often used to simulate natural phenomena such as mountains, marble, fire, and clouds in computer graphics displays. The noise function is described as follows: For every integer location in the (x, y) plane (the *integer lattice*), we define a uniformly distributed, pseudorandom value, v, on the range $[-1,1]$. If x, y is on the integer lattice, we define $noise(x, y) = v_{x, y}$. If x, y is not on the integer lattice, we compute a cubic polynomial interpolation between lattice points using Equation 1 to compute the interpolation coefficient.

$$i = 3x^2 - 2x^3. \qquad (1)$$

This yields a smooth, differentiable function that can then be used to create the surfaces for the stimuli. In order to generate turbulent surfaces, n octaves of this noise are summed, yielding a fractal-like Mónge form surface, or more formally,

$$z(x, y) = \sum_{1}^{n} \frac{noise\ (x, y)}{f^n}. \qquad (2)$$

Figure 3. An example of the stimulus used in Experiment 1. On the left is a gray-scale rendering of the surface with its texture and on the right is a depth map whose units are in degrees of visual angle. Darker regions of the depth map are higher (closer to the observer), lower regions are lighter. The surface was created by summing different scales of a noise function, resulting in a self-similar turbulent surface. This particular surface is the result of the summation of noise that subtended 2°, 1°, and 1/2° of visual angle.

These surfaces differ from true fractals in that they are everywhere continuous, and thus differentiable. Note that by increasing n, and therefore adding more octaves of noise, smaller bumps and dimples result on this type of surface.

In these experiments, three octaves of noise were summed to create the turbulent surface. The base wavelength of the noise function was chosen such that the resulting surface had gross features (bumps and dimples) whose size subtended roughly 2° in visual angle (2 cm). The two additional harmonics added 1° and 1/2° features, respectively. The resulting summation yielded a bumpy surface somewhat reminiscent of a mountainous landscape where the displaced z coordinates of the surface had a range of roughly 10 cm. An example of the stimuli used can be seen in Figure 3.

The Mónge surface was mathematically infinite in extent in both x and y. An arbitrary (x, y) position on the surface was chosen as the center of the test area. Twenty-five probe locations were arranged in a 5 × 5 grid, 5 cm² (5° of visual angle) in size, centered over this fiducial point (see Figure 4). Each location was jittered by a random amount up to 20% of the grid spacing (1 cm) to reduce the possible cues that would be provided by the regular spacing of the grid. For each presentation, a 12 cm² (12° of visual angle) portion was "cut out" from the infinite surface. Additionally, the edges of the surface patch were randomly jittered from presentation to presentation to reduce the possibility of using them as landmarks.

Observers. Observers consisted of five adults, the authors, and one additional observer. All were aware of the purpose of the experiment and had normal or corrected-to-normal vision.

Procedure. This experiment consisted of three different presentation conditions. In all conditions, a square portion of the Mónge surface, 12 cm × 12 cm, was cut out such that one of the 25 probe positions was located in its center. The resulting patch was presented to the observer and the probe was marked with a small
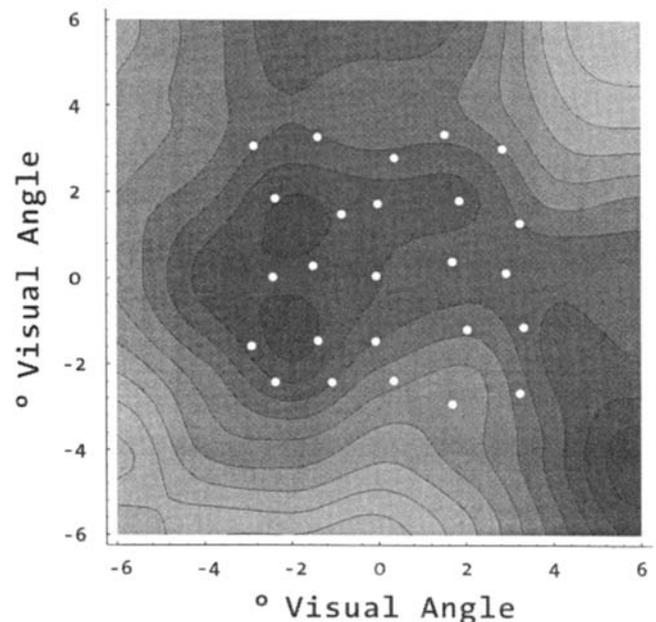
Figure 4. A contour plot of the surface used in Experiment 1 with the probe positions marked with white dots. A fiducial position was selected on the patch and the 25 probe points were distributed in about 5° of visual angle around it.

blue dot that subtended 6' of visual angle. The surface was presented for 5 s, followed by a blank, black screen 5 s in duration. Following the blank screen, the probe location was randomly translated in the (x, y) plane and a new 12 cm^2 surface was cut out. The patch was the same size and general shape (square, barring the random edge perturbations) as the initially presented patch, cut out from the infinite base Monge surface, but the probe location was translated away from the center randomly (see Figure 5). The texture on the surface was also randomly displaced to eliminate the possibility that its patterns could be used for locating the point. The blue dot was randomly repositioned on the surface and its position was made controllable by the mouse. The observer's task was to position the dot in the location shown in the first display.

In the first condition, the initial patch and the adjustment patch were both presented frontoparallel to the observer (that is, with z increasing toward the observer). In the second case, the patch was initially presented frontoparallel but was subsequently slanted 20° and rotated randomly in the viewing plane before the adjustment task. In the third condition the plane was tilted in both the initial presentation and during the adjustment.

Each observer completed three blocks of each condition, for a total of nine blocks. Each block consisted of five adjustments for each probe location, for a total of 15 measurements per probe point per condition.

Results and Discussion

The analysis for each probe position was carried out by calculating an ellipse that surrounded 90% of the judgments made. This resulted in several possible measurements of

error; the primary ones are shown in Figure 6. These errors fall into two primary classes: those connected with the estimation of the probe point location, indicated here by the line l and the direction θ_1, and those concerned with the spread of the estimations, derived from the ellipse that encloses the judgments, such as its orientation, θ_2, and area a. In our analysis, we examined all of these errors and found that the only significant measurement was the spread of the errors calculated by the area (a) of the ellipse. The orientation (θ_1, θ_2) and center point estimation offset (l) errors in this experiment were essentially due to noise.

A graph showing the summary results for each probe point in the experiment is shown in Figure 7. Each probe point location is marked with a dark gray dot; the ellipse enclosing 90% of the estimations is shown in light gray, along with a line connecting the true probe position and the estimated position, defined as the center of the ellipse. These data, in turn, are superimposed upon a depth map of the surface.

As can be seen, the area of the error ellipses is extremely small, less than 5' of visual angle on average in the frontoparallel presentation and adjustment condition. This is particularly impressive because the dot used to mark the probe point is 6' in diameter. For the other two conditions, the error area increases only slightly, with the slanted presentation–slanted adjustment condition performing better than the frontoparallel presentation–slanted adjustment condition. In the conditions where the surface was presented
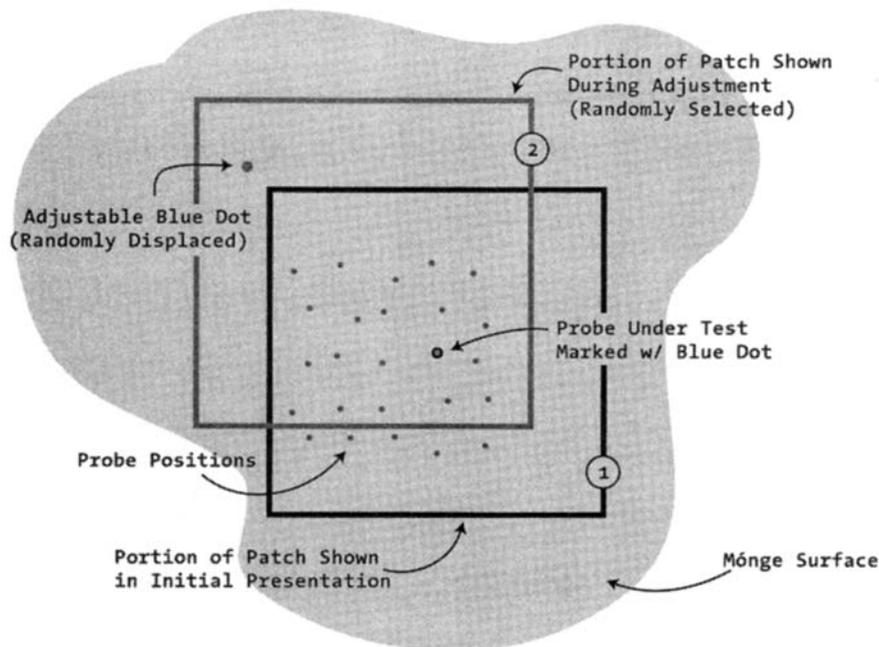
Figure 5. In this diagram, the light gray area represents the infinite Monge surface from which the two test patches are cut for each trial. The potential probe positions are shown in medium gray, and the probe position under test is shown ringed in black. Initially, a section of patch 12 cm × 12 cm is cut out, centered on the probe point under test (Patch 1). For the adjustment portion of the task, an arbitrary patch of the same size is cut out that contains the test point (Patch 2). The dot that was used to mark the probe point in the initial presentation is randomly displaced and made controllable by the mouse.
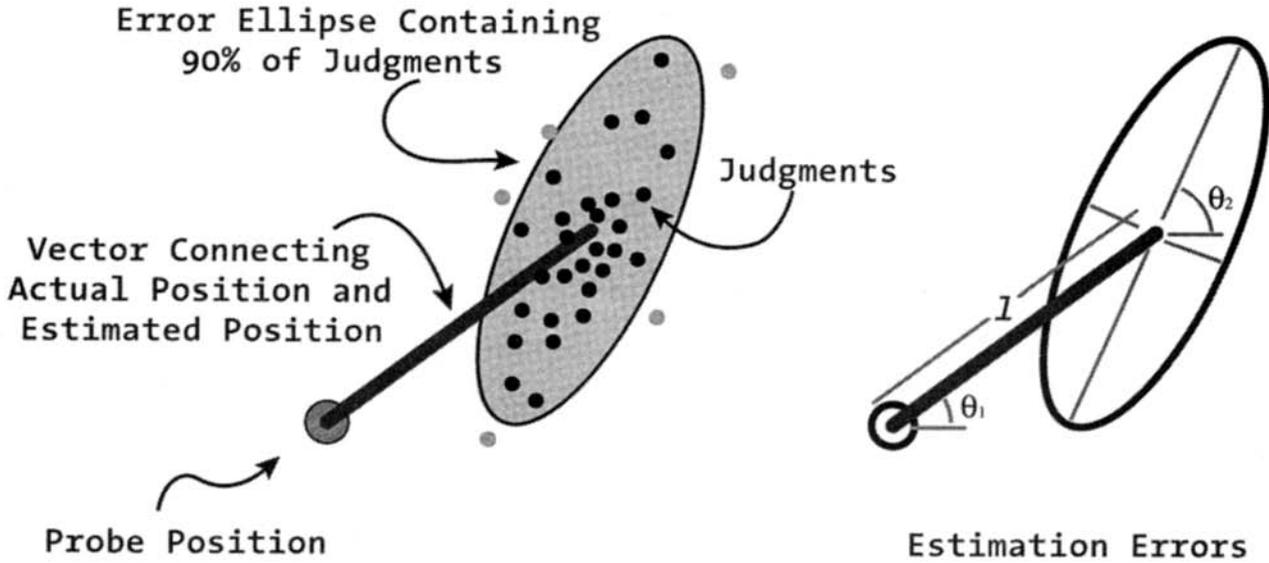
Figure 6. How errors were calculated for Experiment 1. On the left is a diagram showing a probe point and an ellipse fit to 90% of the judgments for that probe. As can be seen on the right, there are several types of error that can be measured. These errors fall into two classes, those connected with the estimation of the probe point proper, indicated here by the line l and the direction θ_1, and those concerned with the spread of the estimations that can be derived from the ellipse, such as the direction of the errors θ_2 and the size and aspect ratio of the ellipse.

slanted away from the observer for adjustment (the second and third), the errors were adjusted by projecting them from the surface into the image plane. A summary graph of these results is presented in Figure 8. As can be seen, there is little difference in the average performance between the three conditions. The lack of effect for the different presentation and adjustment conditions leads us to believe that, at least for the limited orientation differences utilized in the experiment, the perceived identities of individual surface points appear to be viewpoint invariant.

We had initially hypothesized that the errors would be correlated with the surface curvature near the probe points, because locations with higher curvature provide more surface relief and therefore provide a visual anchor or landmark. In

order to test this hypothesis, we can compare the magnitude of the errors with the surrounding region's curvature. Surfaces were constructed by computing the mean error at each probe location and interpolating a three-dimensional surface of the form $(x, y, error)$ that fit these errors such that $error = f(x, y)$. These were further represented as contour plots to help us visualize the structure of the errors in the three conditions. Figure 9 shows how each of these plots relates to the larger probe surface. The probe points only occupy the central part of the surface and therefore the error plots correspond to this area. The error plots for each of the three conditions are shown in Figure 10. Each of these graphs shows a contour plot of a surface fit to the error areas for each of the 25 probe points. That is, areas of high error are
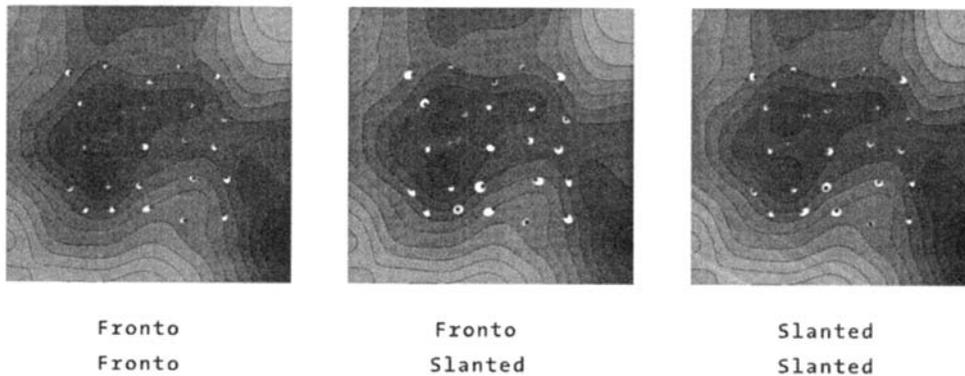

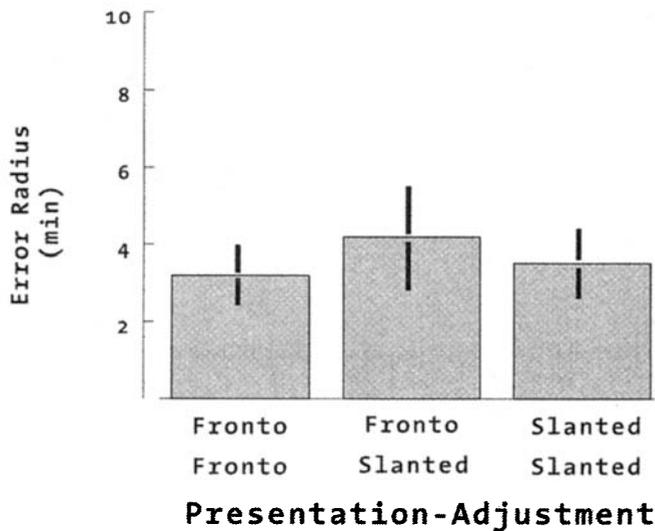|		
Fronto	Fronto	Slanted
Fronto	Slanted	Slanted

Figure 7. The depth map of the surface used in Experiment 1, with ellipses showing the adjustment error for each probe position (across all observers). Note that the ellipses are extremely small, demonstrating that the observers were very accurate in performing this task.

Figure 8. Summary results from Experiment 1 showing the average error over all 25 ellipses plotted against the three presentation and adjustment conditions. The error bars represent ±1 *SD* (averaged across observers). Observers were very competent at this task, as is shown by the extremely small amount of error, which was approximately 4.25' of visual angle in the worst (fronto presentation, slanted adjustment) case. Also note that there was no significant difference between the three conditions.

peaks of the surface, whereas areas of low error are troughs. If the errors at each probe position were independent, then we would expect to see many troughs and peaks, each corresponding to probe positions on the surface. A smoother error surface indicates that there is some similarity in the errors from point to point, and this is what we obtained. The
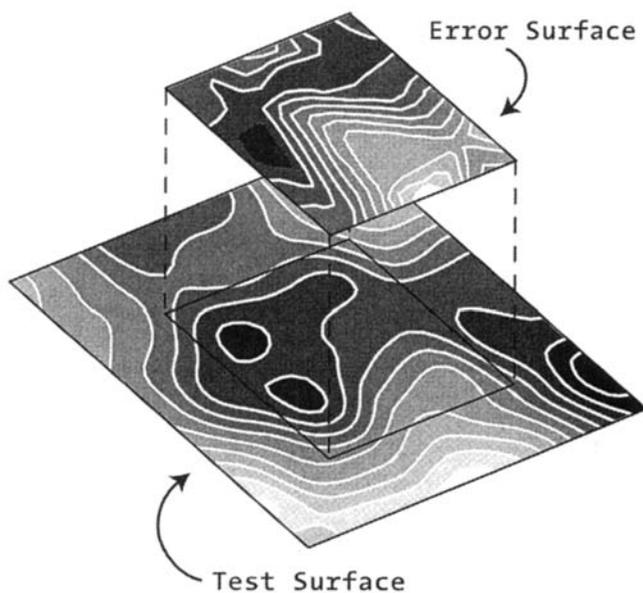
Figure 9. Correspondence between the probe surface and the error surface. The probe points only occupy the central part of the probe surface and therefore the contour plots of the errors correspond to the errors in this central region.

graphs shown in Figure 10 are the error surfaces fit to the error ellipse areas of the central ±2.5° probe region. The area around the "twin peaks" portion of the test surface—see Figure 3 at $(-2°, ±1°)$—has the lowest error and the smooth, steep slope area has the highest. This suggests that surface locations in areas with more structure are easier to localize than smoother, content-free areas. Also notice that the general structure of the errors is the same across all three conditions. This is further evidence that the perceived identities of points on this surface are viewpoint invariant.

The overall structure or change of these errors may depend in part on the curvature in the area surrounding the probe point. A simple thought experiment demonstrates this intuitively. The apex of a cone is easy to locate, but a point on a completely smooth plane or a sphere would be more difficult or impossible if there were no contextual information available. To apply this logic to our experiment, we can compare the amount of error at each point on the surface directly to its underlying differential geometry, in this case its curvatures. In Figure 11 we show the error as a function of the two principal curvatures, κ_1 and κ_2, at a dense sampling of points on the surface in the probe region. As with the error surfaces shown above, a surface of the form $error = f(\kappa_{min}, \kappa_{max})$ was interpolated from the errors and a contour plot was created of the space $(\kappa_{min}, \kappa_{max}, error)$. In this plot, the curvatures are represented on the two axes and the amount of error is represented in the contour plot, white areas representing high error and black areas representing low error. As was suspected, the error was higher in areas where one of the curvatures was close to zero. Looking back at our thought experiment, we see that this is exactly what we would expect—it is difficult to uniquely identify a point along a straight edge (i.e., a Gaussian curvature $\cong 0$). In further support, we see that error was highest when both curvatures were close to zero. In this case, the surface is a relatively flat plane and, again, this result makes sense in terms of our original hypothesis. Finally, when both curvatures were different from zero (i.e., a nonzero Gaussian curvature) the errors were the lowest. This result demonstrates a connection between the surface curvature and our ability to identify a feature on this surface.

We further analyzed the data against other surface properties such as depth and orientation and found no significant effects for any of the error measurements we obtained.

Experiment 2

In describing their subjective impressions while performing this task, all of the observers agreed that there were certain salient structures on the surface, such as hills, valleys, and ridges, that they had used as reference points for localizing the position of the probe dot on each trial. We wondered, therefore, if performance might be impaired by altering the structural complexity of the surface such that the number of potential landmarks would be reduced. Experiment 2 was designed to address this issue.

Method

*Apparatus.* The apparatus and setup were the same as in Experiment 1.
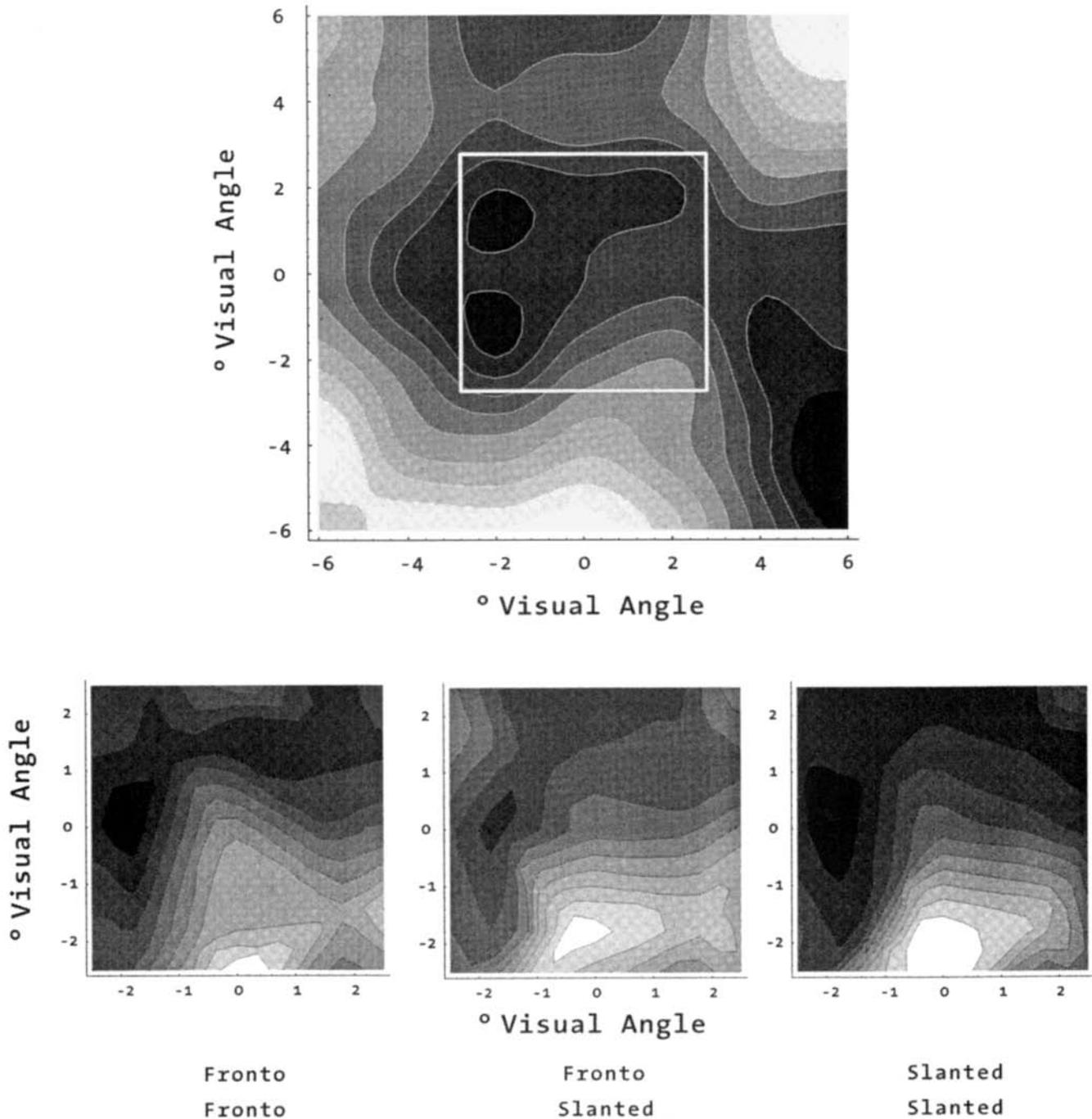
Figure 10. Error surfaces constructed from the estimates from Experiment 1. The upper contour plot shows the probe surface with the area represented by the error plots bordered in white. In the error plots, the white areas represent high error, and the black areas represent lower error rates. The area around the "twin peaks" in the upper left of the probe area has a lower error than the steep slope of the lower right of the probe area. Also, the structure of the error is reasonably consistent across the three conditions.

Stimuli. The stimuli used were the same class of surfaces used in Experiment 1. In this experiment, only the base wavelength of the noise function was used, resulting in features (bumps and dimples) whose gross size subtended 2° in visual angle. This resulted in a much smoother surface with much less structure than in the previous experiment. An example of the stimuli used is shown in Figure 12.

The same fiducial (x, y) position was chosen as the center of the test patch as in Experiment 1 and the set of 25 probe points was the same as well.
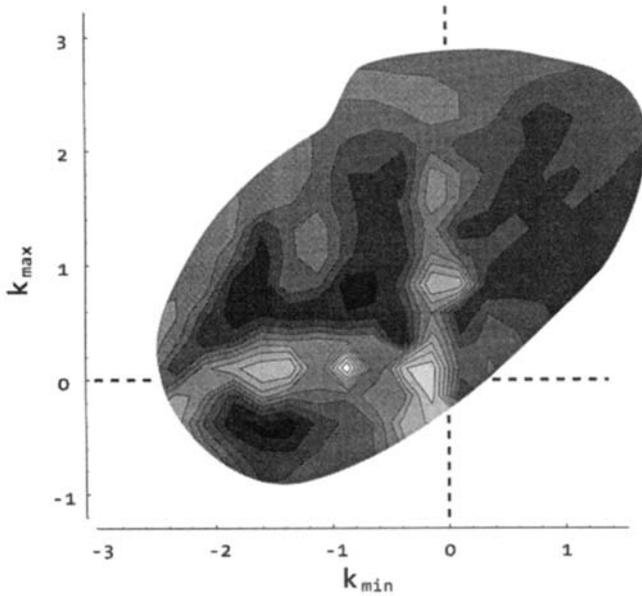
Figure 11.   Error as a function of surface curvature from Experiment 1. Black represents low error, and white represents high error. The area is masked off to show the range of curvatures found in the probe region. Areas close to the zeros (Gaussian curvature of 0) show the highest amount of error, especially where both curvatures are zero. The areas along the zeros, which show moderate-to-high error, are areas where along one of the principal directions surface is flat. Areas of higher absolute Gaussian curvature have the lowest error.

Observers.   Observers consisted of 4 of the 5 participants from Experiment 1.

Procedure.   The experiment consisted of two different presentation conditions. Each observer completed three blocks in each condition for a total of six blocks. Each block consisted of five adjustments for each probe location for a total of 15 measurements per probe point per condition. The presentation methodology was the same as in Experiment 1. The results for the three presentation–adjustment conditions in Experiment 1 were essentially the same across all conditions. Therefore, only the frontoparallel presentation–frontoparallel adjustment and the frontoparallel presentation–slanted adjustment conditions were used in this experiment.

Results and Discussion

The summary results for this experiment are shown in Figure 13. Each probe point's location is marked with a dark gray dot. The ellipse enclosing 90% of the estimations is shown in light gray, along with a line connecting the true probe position and the estimated position, defined as the center of the ellipse. These data are superimposed onto a depth map of the surface.

As with Experiment 1 there was no significant error found in the orientation or position of the error ellipses; the only significant error was in the spread of the estimations, represented as the area of the error ellipses. As can be seen, the average errors have increased significantly over the errors found in Experiment 1. The relative magnitude of the errors appears to increase as one moves across the saddle area on the right hand side of the stimulus. The average error across all 25 probes compared with those of Experiment 1 is shown in Figure 14. The average error for Experiment 2 has risen to roughly 9' of visual angle, double that of Experi-
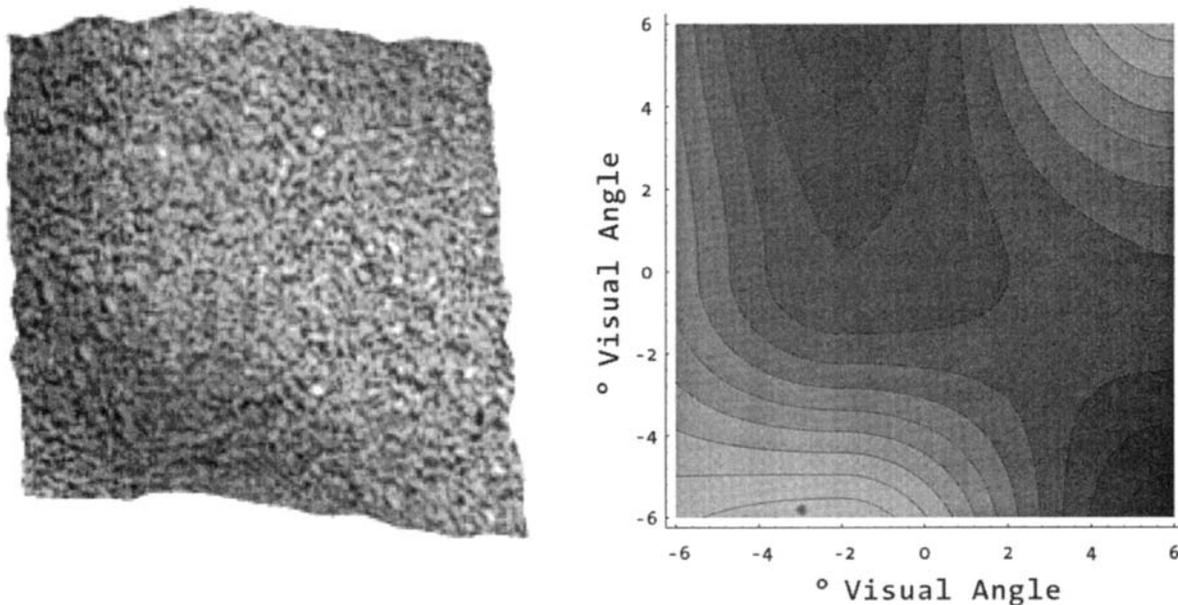
Figure 12.   An example of the stimulus used in Experiment 2. On the left is a gray-scale rendering of the surface with its texture and on the right is a depth map whose units are in degrees of visual angle. Darker regions of the depth map are higher (closer to the observer), lower regions are lighter. The surface differs from Experiment 1 in that it only contains the base 2° wavelength of noise.
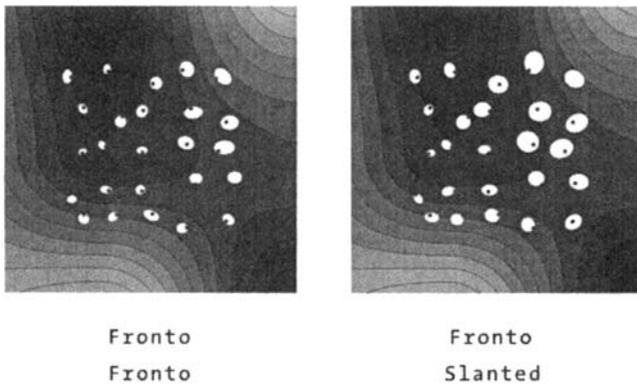
Fronto   Fronto
Fronto   Slanted

Figure 13. The depth map of the surface used in Experiment 2, with ellipses showing the adjustment error for each probe position (across all observers). Note that the ellipses are significantly larger than in Experiment 1. Also, the structure of the errors is slightly more apparent visually, increasing as the probe points approach the saddle-shaped region on the right of the stimulus.

ment 1. Similar to Experiment 1, the presentation–adjustment conditions did not have a significant effect on performance, again suggesting object-centered performance of the adjustment task.

As with Experiment 1, the error surfaces as seen in Figure 15 show a similar overall structure to each other. Again, these plots represent the ±2.5° of visual angle that covers

the probe area, white representing the area of highest error and black the lowest. The increasing of error toward the upper right of the patch can easily be seen in the error surface. As can be seen in the contour plot of the stimulus in Figure 12, the surface flattens out in this area. This is consistent with the results in Experiment 1, which showed greater error in areas of lower curvature.

It should also be noted that there is a significantly smaller range of curvatures available in the surface utilized in Experiment 2. In Experiment 1, curvatures had a range of ±3 cm^{-1}, whereas in Experiment 2 the removal of the higher frequency noise resulted in a range of approximately ±0.5 cm^{-1}. Figure 16 shows the error as a function of the surface curvatures. Areas of low Gaussian curvature again have greater error than those of higher Gaussian curvature. The lack of landmarks, reflected in the low dynamic range of curvatures present on the surface, in conjunction with the reduction of local structure has made it harder to triangulate the position of probes on this smoother surface. These results further reinforce our supposition that the amount of structural information available on the surface effects the ability to identify locations on that surface. Clearly, this is some function of the scale of this information. If we had asked observers to locate probes placed on a flat plane or sphere, the task would have been impossible; likewise, if we were to ask observers to locate points on an extremely rough surface such as a gravel driveway, the task would no doubt have been impossible as well.

Presentation - Adjustment

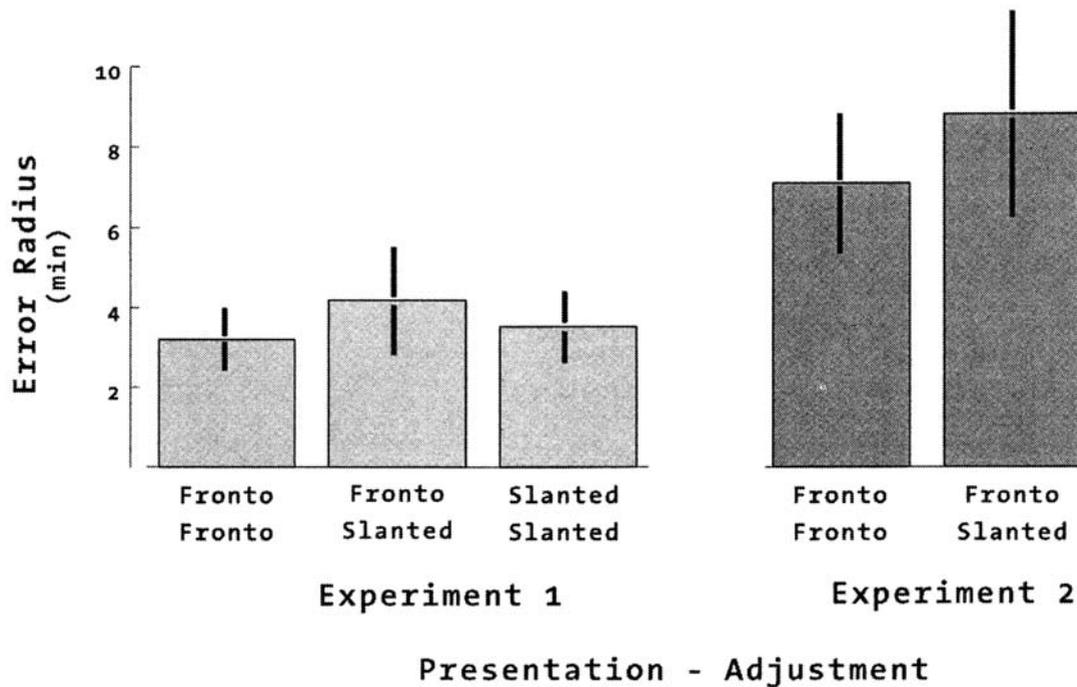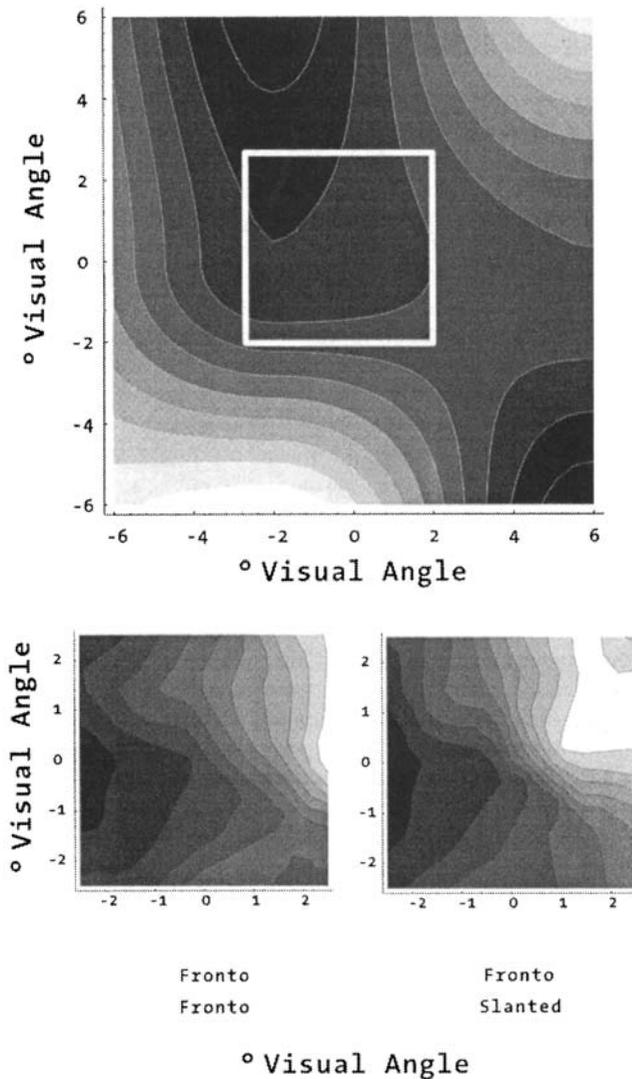*Figure 14.* Summary results from Experiments 1 and 2 showing the average error over all 25 ellipses plotted against presentation and adjustment conditions. The error bars represent ±1 *SD*. Observers in Experiment 2 did significantly worse than those in Experiment 1, where errors were on the order of 9'. As with Experiment 1, the presentation–adjustment conditions did not have a significant effect on performance.

basis of local patterns of surface reflectance, the observers were able to perform this task with a surprising degree of precision. Under optimal conditions, the variance in their settings over multiple trials was just a few minutes of arc.

One important factor that influenced performance on this task was the structural complexity of the depicted surface. The method used to generate our stimuli allowed us to restrict the local variations of surface structure to a particular range of spatial scales. When the structure was restricted to a base wave length of 2°, the average spread of the observers' judgments was approximately 8' of arc within the depicted image. However, when we added additional harmonics to displays, with wavelengths of 1° and 1/2°, respectively, the variability of the observers' judgments was reduced by half to approximately 4' of arc. This finding suggests that the precision with which the position of a surface point can be localized is most likely based on the relative spatial frequency of an object's surface undulations.

One possible hypothesis to consider about the effects of structural complexity is that the precision varies as a fixed proportion of the smallest spatial scale on the surface to which human observers are perceptually sensitive. The present results might appear at first blush to contradict this hypothesis, since the wavelengths used in Experiments 1 and 2 differed by a factor of four, but there was only a twofold difference in the magnitude of the observers' adjustment errors. However, there is other research indicat-

Figure 15. Error surfaces from Experiment 2. The contour plot at the top shows the entire probe surface with the central probe region delineated by a white rectangle. The contour plots at the bottom are the contour plots of the error surface fit to the errors at the 25 probe positions. The white areas show regions of high error, and black areas represent low error.
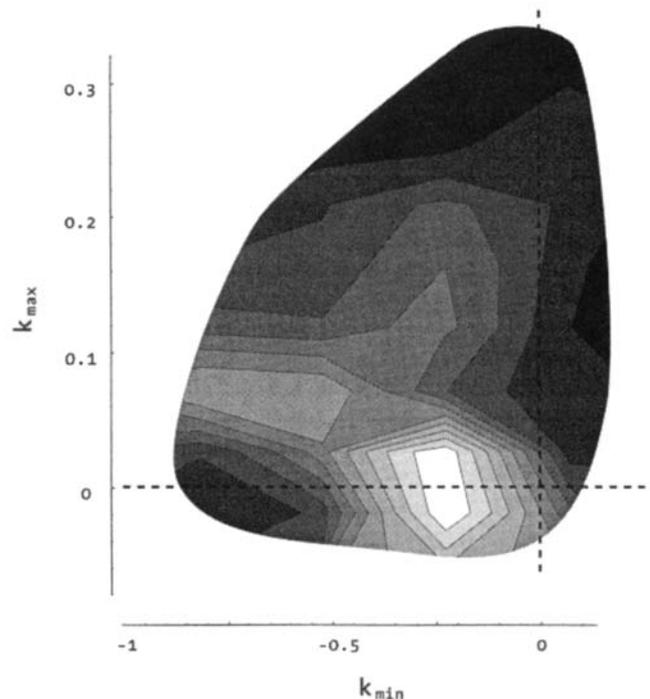
General Discussion

The research described in this article was designed to investigate the abilities of human observers to localize the positions of individual surface points viewed from different orientations. On each trial, a surface defined by shading, texture, and binocular disparity was presented initially from one perspective with a single target location marked by a small dot. Following a brief blank interval, the same surface was presented again from a different perspective, and the observer was required to position an adjustable probe dot to the same location as the target in the first view. Although each presentation of the surface had a different random texture so that the matches could not be performed on the

Figure 16. Error as a function of surface curvature from Experiment 2. Black represents low error and white represents high error. The area is masked off to show the range of curvatures found in the probe region. As with Experiment 1, the areas close to the zeros show the highest amount of error, whereas areas of higher absolute Gaussian curvature have the lowest error.

ing that the human stereo system may be relatively insensitive to the highest frequency surface undulations of 2 cycles/degree that were present in our displays (see Norman, Lappin, & Zucker, 1991; Rogers & Graham, 1982). If so, then the highest perceptible frequencies in Experiments 1 and 2 would have differed by the same proportion as did the variance in the observers' adjustments.

How might it be possible for the human visual system to accurately localize the position of a surface point even though it is viewed from varying orientations? Consider, for example, a typical computational analysis of 3-D structure from optical information, in which surface shape is perceptually represented as a map of local depths, orientations, or curvatures. To determine if two surface probe points viewed from different orientations are the same or different, it would first be necessary to align the surfaces using mental rotation (e.g., Shepard & Cooper, 1986; Shepard & Metzler, 1971; Tarr, 1995).

There are several serious problems with this approach as a theoretical account of human perception. First, there is evidence that the ability of humans to mentally rotate objects about arbitrary axes in 3-D space is quite poor (e.g., see Parsons, 1996), which is perhaps not surprising given that this requires a six-dimensional search in the most general case to find the appropriate transformation parameters (i.e., 3 df for translation and 3 df for rotation). Second, there is additional evidence that humans' perceptual representations of local surface depth, orientation, and curvature are surprisingly imprecise, such that reliable discrimination of these properties requires differences on the order of 10–20%— even under full-cue conditions where surfaces are specified with multiple sources of information including shading, texture, motion, and binocular disparity (McKee, Levi, & Bowne, 1990; Norman, Todd, Perotti, & Tittle, 1996; Phillips & Todd, 1996; Reichel, Todd, & Yilmaz, 1995; Todd & Norman, 1995). Finally, there is also evidence that observers' judgments of depth, orientation, and curvature are systematically distorted and that they do not exhibit viewpoint invariance (Norman et al., 1996; Tittle, Todd, Perotti, & Norman, 1995). It is hard to imagine how mental rotation of such a crude, distorted representation could allow perceptual localization of individual points to within a few minutes of arc.

How then are observers able to perform this task with such a high degree of accuracy? In discussing their subjective impressions, all of the observers agreed that they did not focus their attention on the local surface structure in the immediate neighborhood of the target when making their judgments. Rather, they attended instead to particularly salient landmarks, such as the "top of a bump" or the "edge of a cliff," which they used to triangulate the position of the target when it appeared subsequently in a different orientation.

This strategy is somewhat analogous to how a pirate might hide his treasure on an uncharted island so that he can retrieve it later. To ensure that he can find his treasure again once it is buried, the pirate might begin by exploring the island to find a set of landmarks or features that are perceptually distinct regardless of the context from which

they are viewed. Suppose, for example, that the island contains a single river and an unusual rock formation shaped like a pillar. The pirate might decide to pace the distance between the rock formation and the river and to bury his treasure at the midpoint. He would then be able to find the same spot on a subsequent visit to the island by its spatial relations with respect to these landmarks.

Let us now extend this analogy to the localization of points on smoothly curved surfaces. For a surface point to qualify as a landmark, there are several criteria that need to be satisfied. First, there must be some local property of the point that makes it stand out from its neighbors. Second, that property must be viewpoint invariant. Suppose, for example, that a surface is perceptually represented using a depth map. Within this representation there would be singular points defined by local maxima or minima of depth that could be clearly distinguished from their neighbors, but they would not satisfy the criterion of viewpoint invariance. Similarly, if a surface is represented using an orientation map, there would be singular points defined by local maxima and minima of surface orientation in different directions, but they too would not be viewpoint invariant.

Maxima and minima in curvature are more likely properties for defining landmarks because they are intrinsic aspects of surface structure that do not depend on any particular viewing perspective (cf. Biederman, 1987; Hoffman & Richards, 1985; Richards, Koenderink, & Hoffman, 1987). Moreover, they also satisfy a third criterion that is needed to define a perceptually distinct landmark. In order to be identified using optical information, a singular point on a smoothly curved surface must have a corresponding singular point in the pattern of optical stimulation that is also viewpoint invariant (e.g., Koenderink & van Doorn, 1976; Koenderink & van Doorn, 1980; Koenderink & van Doorn, 1982). To a first approximation, the maxima and minima of surface curvature in any given direction can be optically specified by maxima and minima in the second spatial derivatives of optical properties such as motion, texture, or binocular disparity.

If this hypothesis is correct, then it is likely to be the case that the localization of landmarks defined by maxima or minima of curvature should be significantly easier than the localization of other points whose positions must be determined by their spatial relations with respect to other landmarks. The results of the present experiments have revealed that local curvature in the neighborhood of each probe point is systematically related to the magnitude of observers' adjustment errors, whereas local depth and orientation are not. This provides some support for our argument that curvature is the most perceptually relevant property for localizing surface position. However, because the probe points in the present experiments were positioned at random, the data do not allow a direct test of the hypothesis that maxima and minima of curvature provide a viewpoint invariant frame of reference for localizing the positions of other points when they are presented in different orientations. To address this issue, we are currently planning additional experiments, in which the local surface properties at each probe point will be systematically manipulated

to measure their effects on observers' localization performance.

References

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

DeValois, K. K., Lakshminarayanan, V., Nygaard, R., Schlussel, S., & Sladky, J. (1990). Discrimination of relative spatial position. *Vision Research, 30,* 1649–1660.

Hoffman, D. D., & Richards, W. A. (1985). Parts of recognition. In S. Pinker (Ed.), *Visual cognition* (pp. 65–96). Amsterdam, The Netherlands: Elsevier.

Koenderink, J. J. (1989). *Solid shape.* Cambridge, MA: MIT Press.

Koenderink, J. J., & van Doorn, A. J. (1976). The singularities of the visual mapping. *Biological Cybernetics, 24,* 51–59.

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

Koenderink, J. J., & van Doorn, A. J. (1982). Perception of solid shape and spatial lay-out through photometric invariants. *Cybernetics and Systems Research,* 943–948.

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

McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research, 30,* 137–151.

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

Norman, J. F., Todd, J. T., Perotti, V. J., & Tittle, J. S. (1996). The visual perception of three-dimensional length. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 173–186.

Parsons, L. M. (1996). Inability to reason about an object's rotation orientation using an axis and angle of rotation. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1259–1277.

Peachey, D. R. (1985). Solid texturing of complex surfaces. *Computer Graphics, 19,* 279–286.

Perlin, K. (1985). An image synthesizer. *Computer Graphics, 19,* 287–296.

Phillips, F., & Todd, J. T. (1996). Perception of local three-dimensional shape. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 930–944.

Reichel, F. D., Todd, J. T., & Yilmaz, E. (1995). Visual discrimination of local surface depth and orientation. *Perception & Psychophysics, 57,* 1233–1240.

Richards, W. A., Koenderink, J. J., & Hoffman, D. D. (1987). Inferring three-dimensional shapes from two-dimensional silhouettes. *Journal of the Optical Society of America, 4,* 1168–1175.

Rogers, B. J., & Graham, M. E. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research, 22,* 261–270.

Shepard, R. N., & Cooper, L. A. (1986). Mental images and their transformations. Cambridge, MA: MIT Press.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701–703.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review, 2,* 55–82.

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 663–678.

Todd, J. T., & Norman, J. F. (1995). The visual discrimination of relative surface orientation. *Perception, 24.*