

Foundations of Spatial Vision: From Retinal Images to Perceived Shapes

Joseph S. Lappin and Warren D. Craft
Vanderbilt University

Vision is based on spatial correspondences between physically different structures—in environment, retina, brain, and perception. An examination of the correspondence between environmental surfaces and their retinal images showed that this consists of 2-dimensional 2nd-order differential structure (effectively 4th-order) associated with local surface shape, suggesting that this might be a primitive form of spatial information. Next, experiments on hyperacuties for detecting relative motion and binocular disparity among separated image features showed that spatial positions are visually specified by the surrounding optical pattern rather than by retinal coordinates, minimally affected by random image perturbations produced by 3-D object motions. Retinal image space, therefore, involves 4th-order differential structure. This primitive spatial structure constitutes information about local surface shape.

Vision is a marvelously effective channel of information about the spatiotemporal structure of the environment. It enables rapid perception and recognition of objects, localization of objects relative to the observer and to one another, locomotion through cluttered environments, and coordination of actions with the trajectories of moving objects. How two-dimensional (2-D) retinal images provide such information about the 3-D environment is a long-standing fundamental problem.

Vision involves representations of spatial structure in one physical domain by that in another—involving correspondences between environmental objects, retinal images, neurophysiological signals, perceptual experiences, and behavioral actions. Effective visual function requires reliable mappings between these domains, so they must exist, but they are poorly understood. The general theoretical problem of spatial vision is to identify the spatial properties that are preserved in the mapping of spatiotemporal environmental structure into spatiotemporal structure in the nervous system and subsequent perceptual experience and action.

The specific aim of the present study is to identify the primitive spatial relations that define visual positions of image features and shapes of environmental surfaces. We used visual acuity as an experimental tool to address this problem. Spatial relations be-

tween disconnected surfaces and between the observer and objects were not examined here.

The theoretical ideas are developed as a series of definitions, propositions, and hypotheses. The first proposition concerns the nature of spatial information and how it entails corresponding structures in multiple domains. The next four propositions concern the spatial correspondence between environmental objects and retinal images, and the nature of this correspondence then motivates hypotheses about the spatial primitives for vision. We then discuss alternative hypotheses about the spatial primitives for vision that we tested using experiments on acuity for relative motion and binocular disparity. The results of these experiments offer an understanding of how the spatial structure of retinal images leads to the perception of surface shape.

The Nature of Spatial Information

One basic theme of this study concerns the nature of spatial information. Spatial vision requires reliable and approximately one-to-one spatial correspondences between environmental objects, retinal images, and their representations in the visual nervous system and perception. The spatial information for vision is defined by such mutual correspondences.

This conception of information differs in several respects from that implicit in much of the literature on perception, neuroscience, and cognition. First, the idea that information is defined by a structural correspondence contrasts with the common notion of information as a property of an individual object or event, associated perhaps with its structural complexity or entropy. Although the quantity of information depends on structural complexity, it depends as well on a correlation between the transmitted and received signals. Whether or not something constitutes information at all depends on such a correlation.

Second, information transmission involves correspondence of form rather than transfer of energy. The conception of information as form rather than energy was articulated especially by Wiener (1948, 1954) for understanding communication and control. Communication requires energy, but the information is contained in the form or pattern rather than the energy as such. Indeed, concepts of information, communication, and control are valuable for describ-

Joseph S. Lappin and Warren D. Craft, Department of Psychology, Vanderbilt University.

Warren D. Craft is now at the Department of Mathematics and Computer Science and Department of Psychology, The University of the South.

This article is based on a PhD thesis by Warren D. Craft (1997), *Visual Definition of Spatial Position in Stereoscopic and Moving Patterns*, Vanderbilt University. Support was provided in part by Vision Core Grant P30-EY08126.

This article has significantly benefited from the discussion and suggestions of Warren D. Craft's thesis committee—Randolph Blake, Carolyn Cave, Timothy McNamara, and Alan Peters—and from additional comments by Jonathan Doner, Jan Koenderink, Gordon Logan, and James Todd on an earlier version of this article.

Correspondence concerning this article should be addressed to Joseph S. Lappin, Department of Psychology, Vanderbilt University, 301 Wilson Hall, Nashville, Tennessee 37240-0009. Electronic mail may be sent to joe.lappin@vanderbilt.edu.

ing and quantifying the influence of one physical system on another in cases where this influence is not described by the natural laws governing movements of mass and transfers of energy. Communication and control often involve functional relations between systems that are physically very different from one another.

Vision is just such a case, involving functional relationships among structures and events in at least three very different physical domains: (a) environmental objects and motions; (b) retinal images; and (c) brain, perceptual experience, and action. (The third of these domains involves many separate representations, of course, most of which cannot yet be directly described and distinguished, and so they are included in a single category for present purposes.) Spatial vision is possible only to the extent that mutual correspondences exist among spatial relations in these three domains. Retinal images provide visual information about environmental objects only by virtue of some reliable correspondence or map between object structure and image structure. Similarly, for this image information to be visually effective, its representation in the visual system must be statistically reliable and approximately one-to-one. The reliability of these visual representations may be psychophysically evaluated by the precision of discriminations of spatial relations in the retinal images or the environment.

Figure 1A provides a schematic description of vision as involving one-to-one spatial correspondences between environmental objects, retinal images, and visual representations in the nervous system, perceptual experience, and action. This diagram is contrasted with Figure 1B, in which information-carrying signals are depicted as flowing in time through a series of processes and representations. Figure 1B illustrates the type of diagram typically used in current textbooks on perception, and this scheme is implicit in most contemporary conceptions of perception, cognition, and sensory physiology. Figure 1B does not indicate the nature of the information that flows between successive stages and does not suggest the existence of any particular structural correspondence between these stages. In contrast, Figure 1A focuses on the definition of spatial information rather than its physical transformations.

What is the relation between the present conception of spatial information and Shannon's statistical theory of information and communication (Shannon & Weaver, 1949)? Shannon's statistical theory and the present spatial theory may be seen as two special cases within a more general framework in which information is defined by a correspondence between relational structures in two or more separate domains.¹ In Shannon's theory, the relational structures are merely nominal categorizations in which the elements are arbitrary subsets and simple binary relations (on a pair of elements) are merely *same* or *different*—that is, any two elements are members either of the same set or of different sets. One natural generalization of Shannon's theory is to include structures with stronger relations, such as ordinal or metric. Another related generalization concerns the complexity of the relational structure, involving relations among larger numbers of elements. Metric and topological structures, for example, would be defined by relations among at least three elements. In any case, relations in space and time invite description by structures more complex than the simple categorical structures of Shannon's theory.

The present study is concerned with relational structures that we refer to generically as *spatial structures*. The study of these and

other forms of information is analogous to the study of fundamental measurement theory (Krantz, Luce, Suppes, & Tversky, 1971; Roberts, 1979). Measurement theory is concerned with the representation of qualitative empirical relations by numerical relations, but theories of information concern representations of relational structures in one physical domain by those in another. A principal issue in both cases concerns the conditions under which a specific relational structure in one domain may be represented by that in another—where there is a one-to-one correspondence between relational structures in the two domains.

Definition 1: Spatial Structure

A *spatial structure* consists of (a) a set of texture elements that identify discrete positions within 2-D spatial manifolds, (b) at least ordinal relations among these elements, and, often, (c) motions that displace the relative positions of elements.

By *texture elements* we mean any local mark at a fixed 2-D spatial position on a surface or its image. (The reference frame by which these spatial positions are specified is not yet identified, as that is a principal purpose of this study.) These spatial positions are taken as points in smoothly connected, isotropic, differentiable, 2-D spatial manifolds—associated with environmental surfaces and their retinal images. Accordingly, these spatial structures are assumed to include connectedness and at least ordinal relations on pairs of points; *betweenness* and *collinearity* on triples of points; and, potentially, ordinal relations among separate pairs of points—that is, where the spatial separation between one pair may be *equal to*, *less than*, or *greater than* that of another pair.² The spatial structures pertinent to the present study also include *motions* that displace the positions of points and sets of points. Rigid motions in 3-D space, for example, change the relative positions of environmental objects and the retina, thereby deforming the 2-D retinal images of objects. These same image transformations occur also in stereopsis, in which the two monocular half-images are rotated in three dimensions relative to one another.

The present study investigates a generic family of spatial structures of varying complexity, differing in the differential order and dimensionality of spatial relations at any given point. This family of spatial structures is based on a scheme offered by Koenderink

¹ The concept of *relational structure* is defined as in the literature on the mathematical foundations of measurement (Krantz, Luce, Suppes, & Tversky, 1971; Roberts, 1979). Specifically, a *relational structure* consists of a set of *elements*; *relations* (e.g., less than, equal to, not equal to) among these elements; and *operations* that combine or transform these elements. We use this general notion of relational structure to define the more specific concept of *spatial structure*, defined later.

² The intuitive concept of space often is implicitly metric, involving Euclidean distances, but metric relations have very questionable justification for vision. Metric relations on environmental surfaces would require either embedding in a 3-D space or metric relations that vary with the curvature of the surface. More important, metric relations in the retinal images cannot correspond to those in the environment, so they are neither necessary nor even helpful for perceiving environmental spatial relations. The literature on both human and machine vision often has assumed that the input data are represented in a 2-D Euclidean space, but there is no obvious theoretical or empirical justification for this strong assumption. Arguably, this metric assumption about retinal stimulation has constituted a principal impediment to understanding visual perception.

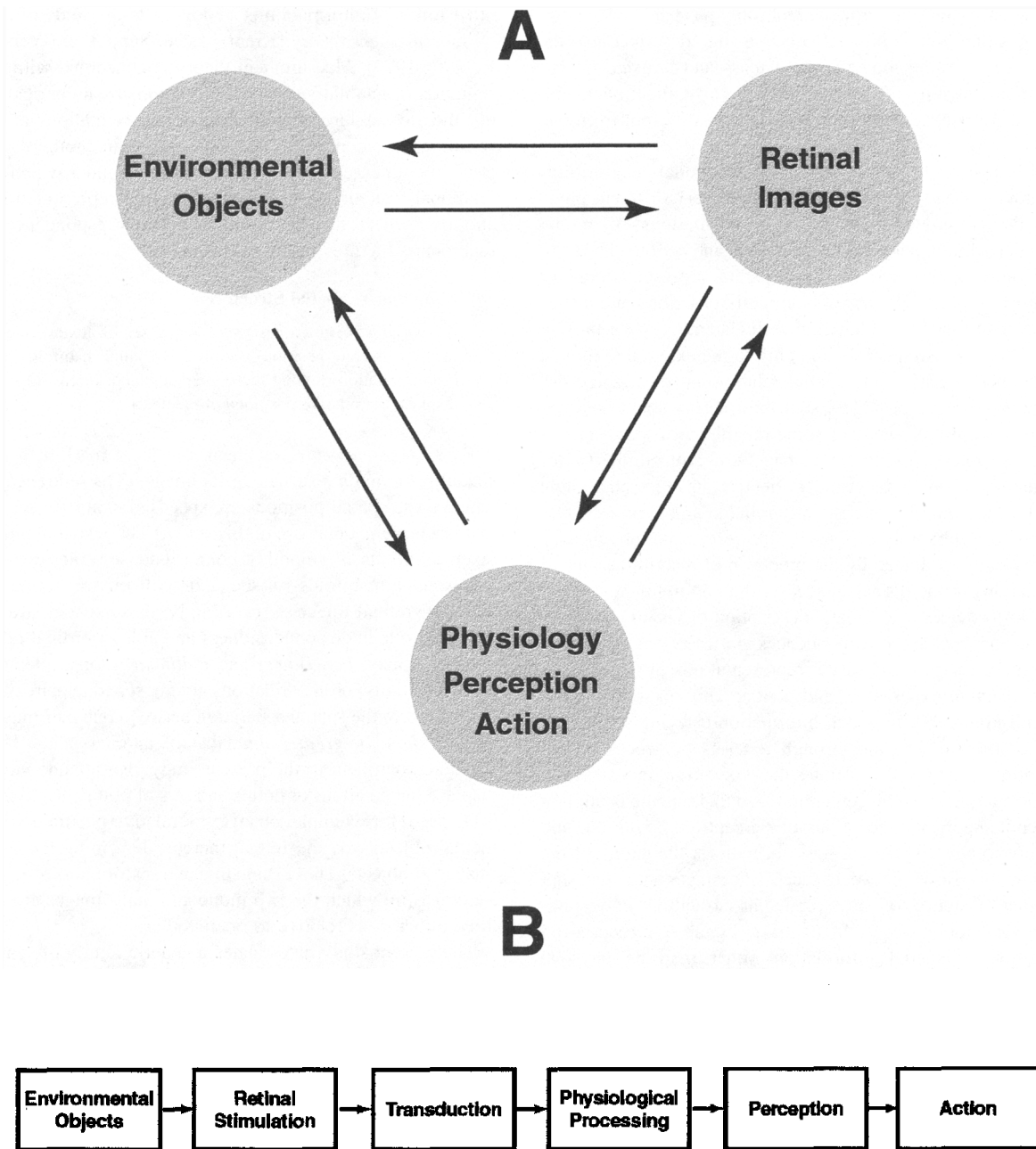


Figure 1. Two contrasting conceptions of spatial information. A: Spatial information is shown as involving one-to-one correspondences between spatial structures in physically different domains. The present study investigated correspondences between the three domains shown in this diagram. (The domain labeled *Physiology, Perception, Action* includes many physically different representations that were not distinguishable in the present study.) B: A common conception of the temporal flow and processing of information in spatial vision. The nature of the information passed through this sequence of representations and processes is not indicated in this diagram. These two diagrams suggest different aims and strategies of research on spatial vision.

and van Doorn (1992b) for describing local image operators, though here we intend no assumptions about the underlying visual operators. Specifically, we examine spatial structures composed of one, two, three, or five points—involving space differentials of zeroth, first, and second order, in one and in two spatial dimen-

sions. Thus, alternative hypotheses about the spatial information for vision are formulated as hypotheses about the simplest relational structure that permits a one-to-one spatial correspondence between environmental objects, retinal images, and visual perception.

The following proposition summarizes these ideas about spatial information for vision:

Proposition 1: *Spatial information consists of corresponding spatial structures in two or more separate physical domains. This spatial correspondence is approximately one-to-one.³ Spatial information for vision involves corresponding spatial structures of (a) environmental objects, (b) retinal images, and (c) visual perception.*

Next, we examine the spatial structures of environmental objects and their retinal images to clarify how images might provide spatial information about objects. Following this examination of the spatial correspondence between objects and images, we use psychophysical experiments to evaluate visual acuities for various forms of spatial structure, with the aim of identifying the primitive forms of spatial structure detected by the visual system.

Retinal Images of Environmental Objects

The mapping from environmental objects onto their images is a projection from a 3-D space onto a 2-D space, and this “loss” of dimensionality often is regarded as a principal difficulty in the inverse problem of inferring environmental structure from image structure. The tractability of this inverse problem depends, however, on assumptions about the spatial structures being mapped. Several different assumptions (often implicit) about these primitive spatial structures are made in the literature on visual perception and computer vision—for example, involving points, 3-D and 2-D separations between points, spatial frequencies, spatial phase structure, edges and contours, tangent planes and image gradients, surface shapes, and higher order derivatives of image properties. The choice of the spatial primitives for describing objects and images has a pivotal influence on the description of image information and a corresponding influence on the conception of visual processes for detecting this information.

Potential primitive spatial structures for vision may be classified by a scheme offered by Koenderink and van Doorn (1992b) for characterizing local image operators—on the basis of the order and dimensionality of spatial derivatives. In this scheme a single region or point is a zero-order structure. A structure of two points or regions would be considered a first-order structure. The spatial parameters of a first-order structure are the length and orientation of the pairwise relations. In general, the number of regions or points in such a structural relation is one greater than the differential order. The number of dimensions and the order of differentiation in a given dimension have conjoint effects on structural complexity, because the number of contrasting regions in the structure equals the product of the number of dimensions and the order of differentiation. Two different types of second-order structures are associated with (a) first-order derivatives in two dimensions and (b) second-order derivatives in one dimension. The former measures gradients in a 2-D space, and the latter measures local symmetry in one dimension. For present purposes, spatial derivatives up to second order and spatial dimensions up to 2-D suffice. We pay particular attention to a “fourth-order” structure, by which we mean a 2-D structure of second-order space-differential relations in the 2-D neighborhood surrounding a given point.⁴

The inverse problem of mapping image structures onto environmental structures generally is more tractable with more complex

image structure. Representations of images as points and their positions within the extrinsic retinal coordinates are ambiguous without specifying the spatial relations among these points. Bishop Berkeley’s “new theory of vision” (Herrnstein & Boring, 1965, pp. 117–125) was founded on the observation that a given point in a retinal image corresponds only to the visual direction of an object point that could exist at infinitely many distances from the observer. These visual directions, moreover, change with movements of the observer’s eyes, head, and body. Thus, spatial positions of image points are uninformative about the 3-D spatial structures from which they arose. Information about environmental structure involves spatial relations, independent of retinal position.

Representing image structure in terms of first-order (pairwise) relations is familiar in visual science with the use of Fourier transforms. Such first-order structure is more informative than zero-order structure about object properties because these pairwise relations are invariant under translations and rotations of retinal position. Nevertheless, the potential correlation between first-order structure in the environment and the image is not preserved under 3-D translations and rotations, which alter the depths and orientations of the object relative to the observer. Hence, the first-order structure is not informative about 3-D separations either within or between separate objects. Image information about 3-D environmental structure is contained in more complex spatial relations, associated with higher order spatial derivatives.

³ The spatial correspondence is qualified as “approximately” one-to-one for several reasons: First, the correspondence is always statistically perturbed by physical noise—by the dispersion of light between the object and its image and by biophysical noise affecting neural spikes and synaptic transmission.

Second, the correspondence is also limited by both the spatial resolution and field of view of the imaging system—that is, by the degrees of freedom in the image or other representation. Koenderink (1990, pp. 32–40) showed how spatial structure can be described independently of spatial scale, using spatial derivatives over regions scaled by a Gaussian blur parameter.

Third, the projective visual map from environmental surfaces onto their images involves singularities—for example, where the image of one part of a surface occludes that of another, at self-occluding and boundary contours, where the viewing direction is tangent to the surface. These singularities occur, however, in isolated infinitely small neighborhoods. Elsewhere, the visual map between the surface and its image is smooth. Thus, a smooth subset that we may call the “visible surface” is isomorphic with its image.

⁴ Spatial derivatives are used here as a generic conception of approximately linear descriptions of local spatial structure, without intending particular visual processes for computing such structure and without presuming particular numerical representations. *First-order structures*, for example, are meant to include all approximately linear representations of local pairwise relations—for example, dipole structures, autocorrelograms, spatial frequency amplitude spectra, center-surround receptive fields, and asymmetric Gabor functions with a single zero crossing. As Koenderink and van Doorn (1992b) pointed out, various orders of differential structure at a given spatial location are algebraically independent.

This characterization of local image structure is also simplified by considering only the spatial and not the temporal structure of images. We assume simply that the spatial structure can be described independently of the temporal structure. Neural mechanisms typically are responsive to both spatial and temporal relations, but the spatial and temporal relations usually are linearly separable (e.g., Adelson & Bergen, 1991; Falzett & Lappin, 1983; Wilson, 1980).

Spatial positions and spatial relations among component texture elements may be represented quite differently depending on the complexity of the primitive spatial structure. For zero-order structures, specifications of spatial positions and relations require an extrinsic reference frame. Spatial positions and relations in the retina, for example, are often thought to be specified anatomically, by "local signs" of retinal position. The intuitive concept of space usually involves an extrinsic a priori framework that is independent of the objects it contains, and this same conception has been applied to vision. Definitions and representations of space do not require such extrinsic reference frames, however.

Space may also be intrinsically defined, by the ordered structure of qualitative relations within and between elements. Relative positions and directions on a smooth surface, for example, may be distinguished by intrinsic coordinates on the surface itself, and two dimensions suffice for these intrinsic surface coordinates. In contrast, 3-D coordinates are required to describe the positions of surface points within the extrinsic space in which the surface is embedded. In spatial structures of two or more points, the spatial position of any given point may be represented by the relative separations and directions among points. First-order spatial structures are invariant under translations over retinal or other extrinsic coordinate frames. Higher order structures are invariant under additional transformations. The distinction between extrinsic and intrinsic reference frames for representing spatial relations has fundamental implications for representing spatial information and for ideas about the visual mechanisms that process this information.

The intrinsic optical structure of retinal images may be described with only weak assumptions about the underlying retinal coordinates—only a homogeneous ordinal structure, satisfying approximate invariance of the spatial derivatives of the image under retinal translations and rotations. Betweenness, for example, is a spatial property definable without reference to an extrinsic coordinate system. Bisection and reflective symmetry are related examples. Another particularly important case involves spatial relations between images—produced by object motion or stereopsis—where one image offers a reference frame for describing spatial differences in the other. Indeed, visual acuity experiments described below demonstrate that spatial positions are visually defined by the intrinsic optical structure of stereoscopic and moving images.

Our discussion of the intrinsic spatial structure of retinal images begins with the following proposition:

Proposition 2: *Retinal images are images of surfaces.*

Optical stimulation of the retina usually is produced by light reflected from surfaces, and the objects of our perception usually consist of such surfaces. From a physical standpoint, this might seem almost trivially obvious, with no strong implications for vision. In traditional conceptions of vision, surfaces have been regarded as complex structures to be inferred from more elementary image properties. More recently, however, the realization has emerged that image structure resembles surface structure.

Gibson (1950, 1966, 1979) emphasized that image structure was associated mainly with environmental surface structure. The theoretical meaning and implications of this idea have been elaborated primarily by Koenderink and van Doorn (e.g., 1975, 1976a,

1976b, 1980, 1992a, 1996; Koenderink, 1987, 1990), using differential geometry. Differential geometry offers a method for describing the intrinsic spatial structure of surfaces and their images.

Surfaces are fundamental to the correspondence between environmental objects and retinal images partly because both surfaces and their images are 2-D structures. Spatial displacements and derivatives over a smooth surface may be described by changes in just two coordinates, and the same is true for images of surfaces. The spatial coordinates used to designate locations on a surface may have almost any structure, provided that the ordinal structure of the positions is specified; neither straight lines or geodesics between points, nor perpendicularity of coordinate directions, nor metric relations are required.

As an illustration, consider positions on a shirt or coat sleeve, perhaps of a wool fabric, with clearly visible texture and folds, curved over one's arm. Positions are marked by variations in texture, thread colors, local irregularities, and folds of the fabric. From the perspective of a very small creature crawling over the fabric, the surface space is 2-D; positions within a larger 3-D environment are largely irrelevant. The same 2-D space is represented in a retinal image of the surface, with spatial positions on the surface represented in its image by optically specified and corresponding local details of texture, color, and shape. These spatial variations in the image of the surface are essentially isomorphic with those on the surface itself. (The rationale for saying that the surface and its image are "essentially" isomorphic is given in Footnote 3.) The mapping between these two spatial descriptions is illustrated in Figure 2.

Metric spatial relations among the component texture elements depend on the reference frame from which they are described. Metric relations within one coordinate frame are related to those in another coordinate frame simply by linear coordinate transformations.

Suppose that a small spatial displacement on the surface is specified in local surface coordinates by a column vector, $d\mathbf{S} = (ds_1, ds_2)$, where ds_1 and ds_2 are the differentials of the displacement in the two surface coordinates. The same spatial displacement described in terms of the retinal coordinates of its image may be given by a corresponding vector $d\mathbf{R} = (dr_1, dr_2)$, where dr_1 and dr_2 specify the displacement in the two retinal coordinates. The relationship between these two descriptions of a spatial displacement can be expressed by the following linear equation:

$$d\mathbf{R} = \mathbf{V} d\mathbf{S}, \quad (1)$$

where \mathbf{V} is a coordinate transformation given by the following matrix:

$$\mathbf{V} = \begin{pmatrix} \frac{\partial r_1}{\partial s_1} & \frac{\partial r_1}{\partial s_2} \\ \frac{\partial r_2}{\partial s_1} & \frac{\partial r_2}{\partial s_2} \end{pmatrix}.$$

Thus, \mathbf{V} describes the projective visual map of the given surface patch from intrinsic surface coordinates to extrinsic retinal coordinates. \mathbf{V} specifies the spatial structure of this image. The inverse transformation, \mathbf{V}^{-1} , which maps the retinal image onto the surface, also exists and is given by a matrix of the same form, with entries $\partial s_i / \partial r_j$:

$$d\mathbf{S} = \mathbf{V}^{-1} d\mathbf{R}. \quad (2)$$

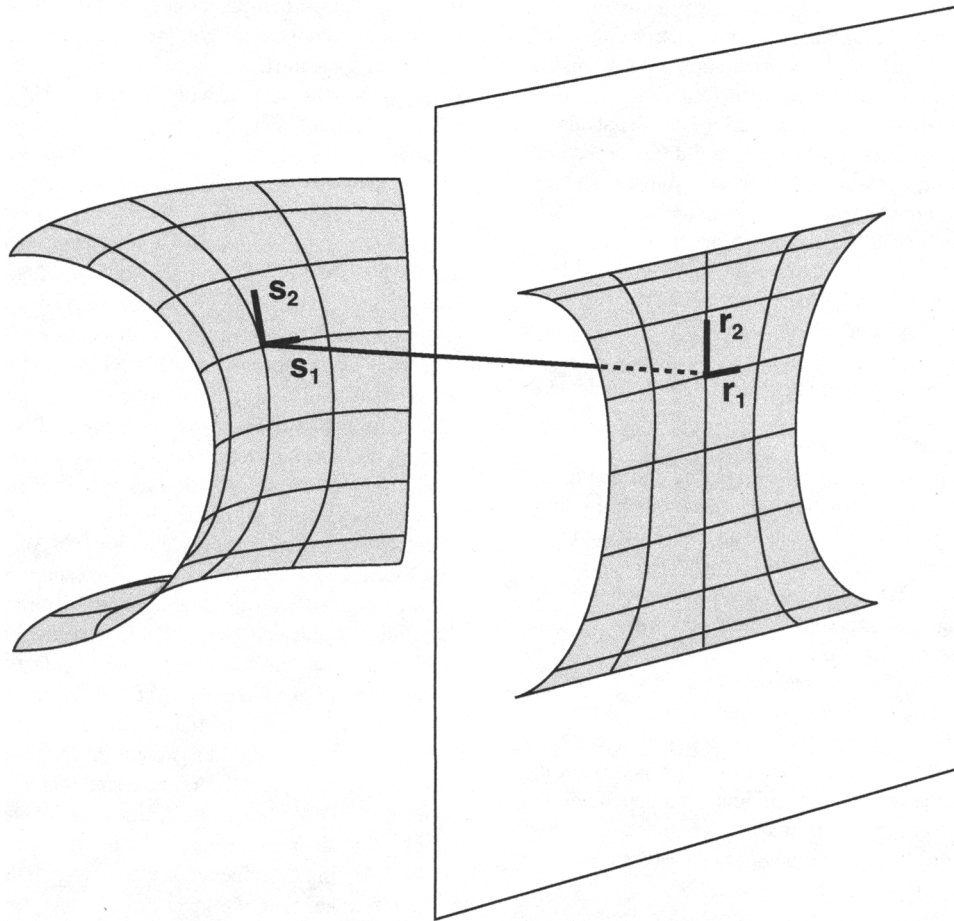


Figure 2. An illustration of the correspondence between spatial displacements in the neighborhood of a visible point on a smooth surface (s_1, s_2) and in the retinal image of the surface (r_1, r_2). The maps between these two spatial structures are smooth, one-to-one, locally linear, and with a well-defined inverse. These maps constitute a *diffeomorphism*.

Equation 1 describes the local image structure. And this local image structure depends only on the local structure of the surface (sufficiently small as to be approximated by a local surface plane), independent of the global surface structure.

As neighboring patches on a curved surface change in 3-D orientation, the linear coefficients of the visual map from the surface onto the image also change. These spatial variations in the local surface orientations, in the visual map from the surface onto the image, and in the image structure are all *smooth*—that is, differentiable. Discontinuities in the surface orientation occur at sharp points, corners, and edges, but these local discontinuities occur at isolated local regions surrounded by smooth variations in orientation. Singularities—where spatial derivatives either vanish or are infinite and undefined—also occur in the visual map and in the image at local contours corresponding to occlusion of one surface region by another part of the same surface or by another object. Boundary contours appear at the outer borders of the image of an opaque solid object, and other occlusion contours may also appear at internal regions of the image where a convex surface region occludes another surface region. Because these singularities occur at restricted local regions surrounded by smooth variations,

the surface, the visual map, and the image may all be regarded as “smooth.” Moreover, the image shapes of these boundary and occlusion contours characterize the surface shape at that position (see Koenderink, 1987; Koenderink & van Doorn, 1976b, 1982).

This correspondence between the differential structures of surfaces and their images is stated more concisely in the following proposition:

Proposition 3: *Surfaces and their images are diffeomorphic.*

The meaning of this statement is that there exists a differentiable map, with a differentiable inverse, between a visible surface and its optical image. In other words, the differential structures of a visible surface and its image are isomorphic; each specifies the other.

What image properties are involved in the differential structure of an image? In stationary images, the diffeomorphism between surfaces and images applies to the correspondence between surface texture and image texture, to spatial relations among visible markers at fixed surface positions. The differential structure of image illuminance, however, is not isomorphic with the surface structure:

Locations and magnitudes of the extrema and gradients of image illuminance depend jointly on (a) the directional distribution of the ambient illumination, (b) the surface orientation relative to the illuminance and image, (c) the bidirectional reflectance distribution function of the particular surface material, and (d) shadows. Therefore, the extrema and gradients of image illuminance do not correspond to fixed surface positions. Nevertheless, image shading constitutes information about surface shape—associated with surface curvature and the second-order image structure. Change in surface orientation—curvature—yields change in image illuminance. Surface locations with large curvature usually give rise to local illuminance extrema. The differential structure of image shading is informative about surface structure, but the correlations are complicated and imperfect (Horn & Brooks, 1989; Johnston & Passmore, 1994a, 1994b; Koenderink & van Doorn, 1980, 1996; Mingolla & Todd, 1986).

A particularly important form of image information about surface structure is provided by spatial relations among multiple images of the same surface, which occur with stereopsis and with image motions produced by rotations of objects in depth. The spatial structure of the image deforms from one such view to another, and these image deformations⁵ are diffeomorphic with the spatial structure of the surface. As we see below, vision is very sensitive to this form of spatial information.

When a surface is viewed binocularly, the two monocular half-images usually differ slightly from one another. Relative to any given reference point, the image positions of other points on the surface differ between the two eyes by amounts that depend on the interocular separation, egocentric distance of the object, and depth variations within the surface. These binocular disparities constitute information about the surface structure.

Very similar monocular image transformations are produced by rotation of a surface in depth. Object rotations entail additional image variations, however, associated with tilt and slant of the rotation axis relative to the image plane, whereas binocular images differ by rotation around a frontal-parallel axis perpendicular to the line between the nodal points of the two eyes. Despite the greater degrees of freedom for the object rotations, the higher order differential structures associated with stereopsis and object rotations are the same.

Vision may well be sensitive to the structural simplicity associated with the smooth maps between environmental surfaces and their disparity and motion fields.

Proposition 4: *The spatial displacements between images of the same surface from neighboring vantage points are diffeomorphic with the visible surface structure.*

The nature of the spatial information provided by these image displacements is contingent on the reference frame by which the relative spatial positions are described. Image displacements associated with stereopsis and object motion constitute vector fields, involving displacements in the relative positions of given points within the two images.⁶

A traditional conception of binocular disparity is based on differences in the retinal positions of a given feature in the two monocular half-images. The relative depths associated with absolute disparities, however, vary with the egocentric distance (roughly proportional to the square of the distance) and direction of the fixation point. Although the binocular vergence angle might

provide some information about egocentric distance, beyond a few meters small changes in vergence angle correspond to very large changes in egocentric distance and even larger changes in the relative depths associated with given disparity values. Observers are quite unreliable in using vergence to estimate viewing distance (Collewijn & Erkelens, 1990; Foley, 1980; Regan, Erkelens, & Collewijn, 1986). Moreover, changes in the distance or direction of the fixation point alter the absolute disparities. Normal viewing entails continuing movements of both eyes, relative to each other and relative to the objects in the visual world, and this fact vitiates depth information that might be provided by absolute disparities.

A more realistic conception of disparity is based on binocular differences in the intrinsic spatial relations among features within the two half-images. Binocular differences in pairwise separations of features, for example, have the important advantage of being invariant under changes in retinal position produced by eye movements. Even so, the information provided by such pairwise (first-order) image properties is greatly diminished by dependence on the relative distances and orientations of objects. If binocular disparities are informative about surface shape, then this information is associated with the higher order differential structure.

Koenderink and van Doorn (1976a) pointed out that the orientation of a local surface patch corresponds to binocular image deformations (*pure shear*) of the 2-D structure of first spatial derivatives, involving binocular differences in the image shapes of surface tangent patches. This image information about the orientation of the surface patch is invariant under expansions (divergence) and 2-D rotations (curl), but it is not invariant under 3-D rotations that change the slant of the surface patch relative to the observer's retinas. Shape involves the spatial changes in local orientation rather than the local orientations per se, and, likewise, image information about shape involves second spatial derivatives of the image.

The local shape at any point on a smooth surface is one of only four alternatives: (a) *planar*; (b) *parabolic*—flat in one direction and curved in the orthogonal direction—cylindrical sections and contours at which the curvature changes sign between convexities and concavities; (c) *elliptic*—curved with the same sign of curvature in two orthogonal directions—hills or valleys, convex or

⁵ We use the term *deformation* to refer to changes in the projected image shape of a local surface patch produced by rotation in three dimensions. By definition, these spatial transformations involve the 2-D image structure, independent of isotropic scale changes. In Koenderink and van Doorn's (1975; Koenderink, 1986) well-known analysis of optic flow patterns, the "def" component involves anisotropic scale changes of opposite sign in orthogonal directions, independent of an isotropic "div" component. This *def* component involves the 2-D first-order image structure, but one can also examine deformations of higher order structure. "Second-order deformations" of the 2-D second-order differential image structure are especially relevant to the present study. Deformations of the first-order structure do not imply deformations of second-order structure, so the first-order changes, involving both isotropic and anisotropic scale changes, may be factored out of the description of second-order deformations.

⁶ Any given optical image point may be taken as a reference position, and the residual disparities in the surrounding image region then constitute a vector field. The spatial positions of these vectors may be defined by reference to the intrinsic image structure as well as by extrinsic retinal coordinates.

concave; or (d) *hyperbolic*—curved with opposite signs of curvature in orthogonal directions—saddle-shaped regions, as at a pass between neighboring peaks. Thus, local surface shape is identified by two independent curvatures. The maximum and minimum curvatures—the *principal curvatures*—provide a convenient description, partly because these are orthogonal.

Surface curvature in a given direction involves a second-order spatial derivative—involving change in orientation of the surface normal relative to displacement over the surface in that direction. The shape of the surface at any point involves the relative magnitudes of the principal curvatures. The ratio of the principal curvatures is a useful numerical scale of shape independent of the magnitude of curvedness (though other reasonable shape scales can also be constructed). That is, local shape involves a relation between two second-order differentials; it is a fourth-order differential property.

Image information about surface shape also involves fourth-order differential structure, involving 2-D structure of the second spatial derivatives around any given point. Figure 3 illustrates these fourth-order image deformations. As may be seen, the four alternative local surface shapes are identifiable by these deformations. A change in the relative collinearity of vertically aligned points represents surface curvature in that direction, and a change in the bilateral symmetry of horizontally aligned points represents surface curvature in that direction. Elliptic surfaces give rise to a uniform direction of displacements relative to the central reference, and hyperbolic surfaces yield opposite displacements in the two principal directions. The bottom row of Figure 3 illustrates the image changes surrounding a central reference point.

The illustrations in Figure 3 have been simplified by examining a special case in which the principal directions of curvature are parallel and perpendicular to a vertical rotation axis. The principal directions of a given surface patch generally are not aligned with the rotation axis, however. For the general case in which the principal directions are not known, the differential image structure must be measured in at least three directions (see Koenderink & van Doorn, 1992a), rather than only two as might be suggested by Figure 3. Optic flow fields generally also involve surface patches and rotation axes that are slanted and tilted relative to one another and relative to the image plane, involving additional interimage displacements besides those shown in Figure 3.

Arbitrary rotation axes generally yield additional changes in first- and second-order image structures. Slant of the rotation axis produces an additional 2-D rotation (curl) component in the interimage vector field. Translation in depth produces isotropic expansion (divergence). Both of these image transformations affect the first-order image structure, but the second-order structure remains invariant. Anisotropic scale transformations, affecting the 2-D image structure, are produced by slant of the surface patch relative to the image plane, but the second-order differential structure is unchanged by any of these affine scale changes. Moreover, the image direction of the projected rotation axis is specified by the second-order directional derivatives of the interimage displacement field, because the second-order derivatives of the displacement vector field are always perpendicular to the projected rotation axis, regardless of its tilt and slant and regardless of the direction of differentiation (Koenderink & van Doorn, 1992a).⁷

Nevertheless, the relative magnitudes of the two principal curvatures and the quantitative surface shape generally cannot be

determined from the interimage displacements of stereoscopic or moving surfaces. The magnitudes of the second-order directional derivatives depend on the surface curvature in a plane normal to the image rather than normal to the surface itself, and this non-normal “sectional curvature” is larger than the normal curvatures that define the intrinsic local surface shape (see Dijkstra, Snoeren, & Gielen, 1994; Koenderink & van Doorn, 1992a).

The qualitative distinction between planar, parabolic, elliptic, and hyperbolic shapes, however, is invariant under changes in orientation of the surface patch and rotation axis. These orientations affect only the scalar magnitude and not the sign of these second-order image measures. Moreover, the relative values of the second-order derivatives of the disparity field are not much affected by the orientation of the surface patch except for large values of curvature and slant (Dijkstra et al., 1994).

Figure 4 illustrates how information about local surface shape is contained in the magnitudes of the second-order directional derivatives, showing the 2-D forms of local image measures around each of several points on four different surfaces. As may be seen, the 2-D structures of these local image deformations specify the qualitative local surface shape, though they also vary with the location of the surface patch. As Koenderink and van Doorn (1992a) pointed out, the forms of these local second-order deformations are related to the *indicatrix of Dupin*—the shape of a small slice through the surface parallel to the tangent plane at a given point. The form of the image deformation, however, corresponds to the projection of Dupin’s indicatrix, and it varies with the relative orientations of the surface tangent plane and the image plane. The spatial information illustrated in Figure 4 is a generalization of that suggested by Figure 3.

Proposition 5 summarizes the preceding ideas:

Proposition 5: *The qualitative local shape at any point on a smooth surface is specified by the 2-D structure of the second-order spatial derivatives of its interimage displacement field.*

This geometric relationship is given more precisely by the following equation (from Koenderink & van Doorn, 1992a). Let D_α represent the directional derivative in a direction α at a given location in the interimage displacement field, ν , of a smooth surface rotated through an angle θ or seen stereoscopically with an angle θ , in radians, between the two monocular view directions, and let $\kappa(\alpha)$ represent the normal curvature of the surface in the direction α . (By definition, $\kappa = 1/\rho$, where ρ is the radius of the best fitting circular arc.) Then the surface curvature is proportional to the second directional derivative of the interimage displacement field, as given by the following equation:

$$D_\alpha D_\alpha \nu = \kappa(\alpha)(\theta/\cos\beta \cos^3\gamma) \mathbf{E}, \quad (3)$$

⁷ “Directional derivatives” arise for spatial derivatives on a multidimensional vector field. Recall that derivatives of vectors are also vectors, with direction as well as magnitude. The directions of these difference vectors depend on the changing directions of the image displacement vectors, rather than the direction of differentiation. For the second-order directional derivatives for binocular disparity and object rotation, the directional components of these directional derivatives are all perpendicular to the projected rotation axis. The magnitudes of these directional derivatives depend on the direction of differentiation and on the curvature of the surface in that direction.

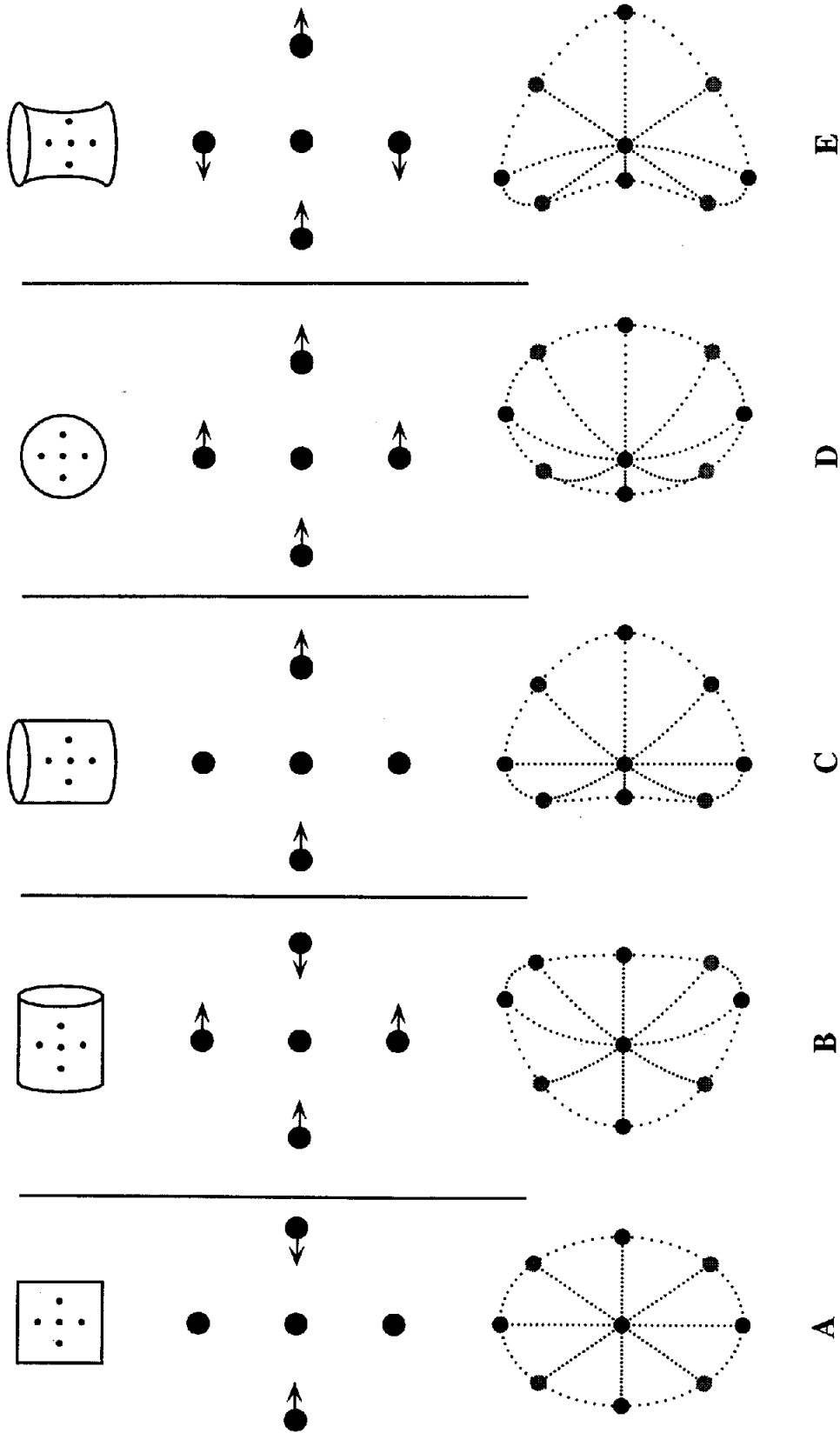


Figure 3. A schematic illustration of the qualitative second-order differential structure of the two-dimensional neighborhood of a given point in the interimage displacement fields associated with stereoscopic disparity or rotation in depth for each of the four possible local shapes of a smooth surface. A: A planar (flat) surface patch. B and C: Parabolic (cylindrical) patches in two different orientations. D: An elliptic (convex or concave) patch. E: A hyperbolic (saddle-shaped) patch. The central point in each diagram is a reference point for describing the surrounding image displacements. (These diagrams were suggested by Steven Tschantz, Department of Mathematics, Vanderbilt University.) Adapted from *Vision Research*, 37, J. S. Lappin and W. D. Craft, "Definition and Detection of Binocular Disparity," p. 2955, Copyright 1997, with permission from Elsevier Science.

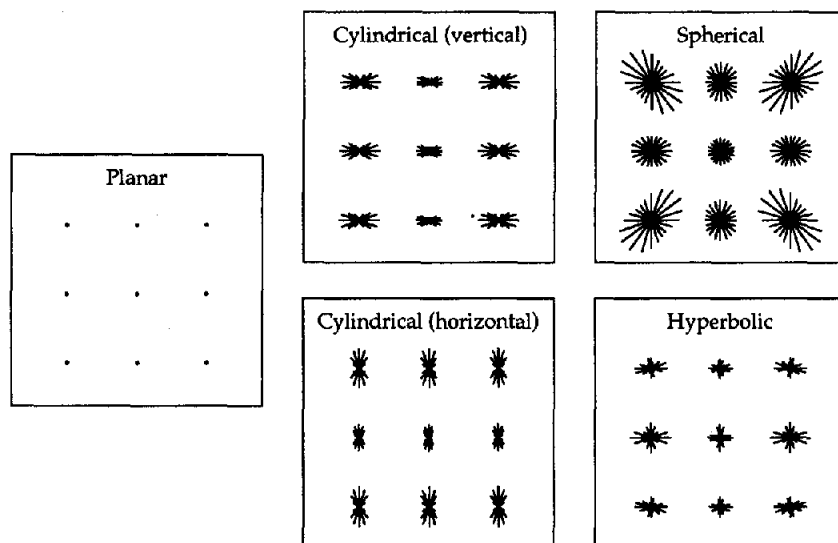


Figure 4. The structure of the second-order directional derivatives of the disparity vector fields at nine points on each of four different surfaces. The axis of rotation was tilted ($\tau = 112.5^\circ$) and slanted ($\sigma = 67.5^\circ$) relative to the image plane. These plots show the signed magnitudes but not the directions of these derivatives of vectors. The directions of these vector derivatives are all in the “epipolar” direction—perpendicular to the projection of the rotation axis. The directions of these magnitudes are the directions of differentiation. For the hyperbolic surface (bottom right), lobes in the two perpendicular directions have opposite signs, but for the other surfaces, the signs are the same in all directions.

where \mathbf{E} is a unit vector in the epipolar direction (perpendicular to the image of the rotation axis), and β and γ are angles that depend on the slant and the tilt of the surface. (Specifically, let B be the fiducial plane defined by the visual direction and the direction α . Then β is the complement of the angle between the surface tangent plane and B , and γ is the complement of the angle between the visual direction and the line of intersection between B and the surface tangent plane.)⁸ For the special case in which the visual direction is normal to the surface, then β and γ are both 0 and the denominator is 1, but otherwise the denominator is less than 1, and the factor in parentheses is larger than θ .

Equation 3 gives the image differential in a single direction, α , but information about the local surface shape involves the 2-D structure of these values in varying directions. Notice that ratios of the form $\pm|D_\alpha D_\alpha v|/\pm|D_\mu D_\mu v|$, where α and μ are two different image directions, are scale-free (dimensionless) measures of image structure. These image measures are also independent of the rotation angle θ . For the special case in which the visual direction is normal to the surface, this ratio is equivalent to $\kappa(\alpha)/\kappa(\mu)$, provided that $\kappa(\mu) \neq 0$, and if $\pm|D_{\max} D_{\max}|$ and $\pm|D_{\min} D_{\min}|$ are the maximum and minimum signed magnitudes of the second-order directional derivatives, then their ratio provides a scale-free index of the local surface shape. To accommodate the case where $|\kappa_{\min}| = 0$, a better shape index is given by

$$\arctan [(\kappa_{\max} + \kappa_{\min})/(\kappa_{\max} - \kappa_{\min})],$$

and a corresponding image estimator may be formed from the second-order directional derivatives of the image (see Koenderink, 1990, pp. 319–324).

In general, the surface slant and the angles β and γ are not zero; the 2-D structure of second-order directional derivatives varies

with slant of the local surface tangent plane. This dependence is illustrated in Figure 4, in which the local shape is the same over the whole surface, but the local differential structure varies among these nine locations that differ in slant of the local patch. Nevertheless, as Figure 4 illustrates, these second-order directional derivatives clearly distinguish among the four categories of local surface shape. The denominator of the parenthetical factor in Equation 3 is always positive, so the sign of the quantity is unaffected. Moreover, the slant angle and direction of tilt are both specified by the first-order structure of the directional derivatives of the displacement field (see Koenderink & van Doorn, 1992a). Thus, the simple relationship given in Equation 3 describes information about the local surface shape.

One last question concerns perspective, involving nonlinear effects of egocentric distance on the mapping from the surface to the image. Perspective may be less problematic than one might first expect, however: Perspective affects the global but not the local image structure of small surface patches. Regardless of perspective, the map from local surface coordinates to local image coordinates is linear, as in Equation 1. The preceding ideas about

⁸ The angles β and γ may be defined in terms of the slant σ and tilt τ of the surface by the following equations (from Koenderink & van Doorn, 1992a):

$$\begin{aligned} \cos\beta &= \cos\sigma[1 + \tan^2\sigma \cos^2(\alpha - \tau)]^{1/2} \\ \cos\gamma &= 1/[1 + \tan^2\sigma \cos^2(\alpha - \tau)]^{1/2}. \end{aligned}$$

The relation between the intrinsic normal curvature of the surface and the curvature in the fiducial plane B varies with the angle β : $\kappa_n(\alpha) = \cos\beta \kappa_s(\alpha)$, where κ_n and κ_s are the curvatures in the normal plane and in the nonnormal fiducial section, respectively.

the spatial structure of surfaces and their images are independent of assumptions about retinal coordinates.

Spatial Input to Vision: Methodological Rationale

The preceding development concerns spatial correspondence between environmental surfaces and their images, but not vision as such. How is this spatial information detected and represented by visual mechanisms?

This is a psychophysical question requiring psychophysical experiments. Presently available physiological methods do not reveal how spatial relations are represented within and among large numbers of neurons in multiple brain areas.

The present investigation used acuity to evaluate the visually defined spatial structure of stimulation. Specifically, we evaluated visual acuity for the position of a given feature in images undergoing various groups of spatial transformations. If acuity can be maintained under a given group of transformations, then the position must be defined by spatial relations that remain invariant under these transformations. That is, spatial positions must be visually specified relative to some reference frame, involving either the extrinsic structure of the retinal mosaic or the intrinsic structure of the image. These reference frames can be distinguished by the transformations under which they remain invariant.

Consider, for example, a stereoacuity experiment on the detectability of binocular disparity of a given target point. If spatial positions were extrinsically specified by reference to the retinal mosaic, then independent random translations of the two half-images would constitute positional noise, and stereoacuity would necessarily suffer. On the other hand, if spatial positions were visually defined in relation to the intrinsic image structure, then spatial positions within the two half-images would not be altered by binocularly uncorrelated image translations. Various groups of image transformations may have different effects on stereoacuity, depending on the particular spatial relations that define visual positions.

Visual Acuity and Hyperacuity

Definition 2: Visual Acuity

Visual acuity refers to the least discriminable difference in spatial position. Acuity may be measured by the standard deviation of spatial positions that are not discriminably different from a given position.

Acuity was measured in the present experiments by the method of adjustment. Observers used a joystick to adjust the position of a given line or point so that it satisfied some specified relation to neighboring features—so that it bisected a spatial interval; so that it was coplanar in depth; or, in relative motion experiments, so that it was stationary relative to the neighboring features. Acuities were measured by the standard deviations of these adjustments. These standard deviations estimate the random noise in spatial information.

A convenient generalized form of this acuity measure is the Weber fraction—where the standard deviation is divided by the separation between neighboring features. The Weber fraction is a scale-free measure of the precision of visual information. It permits comparisons of performance for stimulus patterns with different spatial size, different spatial configurations, different dis-

crimination tasks, and even different sensory modalities. Whether or not obtained standard deviations actually are proportional to the feature separations is a separate empirical question; constancy of the Weber fraction is not assumed in using it to describe performance.

Definition 3: Hyperacuity

Hyperacuity refers to visual discriminations of differences in spatial position significantly below the resolution limits predicted by the point-spread function of the eye and below thresholds obtained in resolution acuity tasks.

Hyperacuity is a term coined by Westheimer (1975) to describe the striking superiority of visual acuity in tasks requiring discriminations of relative spatial position, as compared with that obtained in tasks requiring spatial resolution. Such visual resolution tasks include (a) two-point resolution, requiring a discrimination between one and two points of light, and (b) high spatial frequency. Acuity thresholds obtained in such resolution tasks usually are a little less than 1 arcmin, consistent with what would be expected on the basis of the optical point-spread function of the eye, optical diffraction, or the separation between cones in central fovea.

In contrast, hyperacuity thresholds may be much smaller, from about 1–20 arcsec (e.g., Berry, 1948; Klein & Levi, 1985; Westheimer, 1977, 1979b; Westheimer & McKee, 1978, 1979). Such hyperacuities have been obtained by many investigators for many tasks including (a) vernier alignment, (b) bisection of the distance between parallel lines separated by several arcmin, (c) motion detection, and (d) stereoacuity for binocular disparity (see Westheimer, 1979b).

The fact that acuity in any visual task could surpass physical resolution limits may seem paradoxical, but the explanation is that these hyperacuities involve differences in spatial position and not resolution of overlapping luminance distributions. The smallest area of retinal illumination from an easily visible point of light has a diameter of several arcmin, stimulating a dozen or so neighboring photoreceptors, but the intensity is differentially distributed over the receptors, providing precise estimates of the centroid of the distribution. Small changes in location of the centroid produce widespread changes in the relative stimulation of neighboring photoreceptors and neurons, and such shifts in the distributions permit fine-grained discriminations of spatial position. Indeed, differences in relative positions of two points are more easily discriminated if their retinal illuminances do not overlap extensively. Westheimer (1977, 1979b), Klein and Levi (1985), Geisler (1989), and others have provided clear and plausible explanations for hyperacuities of a few seconds of arc.

Hyperacuities usually have been attributed to local mechanisms responsive to differences in stimulation of individual receptive fields. Local mechanisms seem to be involved in hyperacuities for bisection and vernier alignment of stationary patterns, where the component features must be separated by no more than a few minutes of arc. Local motion-sensitive mechanisms also seem to be implicated in detecting motion. Similarly, stereoacuity for binocular disparity has been attributed to cortical neurons sensitive to binocular retinal positions.

This simple local model of stereoscopic acuity for disparities in absolute retinal positions is invalidated, however, by several experimental results (for a review, see Lappin & Craft, 1997), in-

cluding the following: First, stereoscopic hyperacuties are obtained for differences in the depths of features separated by several degrees, with Weber fractions around 0.25% of the separation between features (see Experiment 1 below). Second, stereoacuties remain robust under randomly changing dichoptic image translations uncorrelated between the two monocular half-images (Lappin & Craft, 1997, and Experiment 1 below). Such binocularly uncorrelated image translations would produce noisy variations in the stimulation of binocular neurons sensitive to absolute retinal disparities, but this retinal position noise has little effect on stereoacuity so long as spatial relations among the features in each monocular half-image remain constant. Thus, stereoacuity seems to involve spatial relations among features within each half-image.

The neural mechanisms underlying stereoacuity are not yet understood, but evidently they involve somewhat widespread spatial relationships within each monocular half-image. The extensive spatial range of the mechanisms implicated in stereoacuity turns out to be characteristic also of hyperacuties for motion of one feature relative to its neighbors, as shown in the experiments presented here. We refer to these spatial sensitivities as multilocal—to emphasize that the spatial range of the underlying mechanisms is more extensive than that of the local mechanisms implicated in hyperacuties for stationary nonstereoscopic patterns.

The following two propositions express the hypothesis that visual acuties involve sensitivities to at least four qualitatively different types of spatial relations:

Proposition 6: *Four different types of visual acuity may be distinguished: (a) resolution, (b) discrimination of local relative position, (c) discrimination of multilocal structure in stationary nonstereoscopic patterns, and (d) discrimination of multilocal differences in stereoscopic depth or relative motion. These four types of acuity depend on different spatial characteristics of stimulation and, presumably, different visual mechanisms.*

Proposition 7: *Hyperacuity may be obtained in discriminations of local relative positions and of multilocal differences in stereoscopic depth or relative motion.*

The hypothesis that hyperacuties for binocular disparity and relative motion involve multilocal spatial sensitivities different from those for stationary nonstereoscopic patterns is supported by several recent experiments (e.g., Lappin & Ahlström, 1996; Lappin & Craft, 1997; McKee, Welch, Taylor, & Bowne, 1990) and is further tested in the experiments presented here. These multilocal hyperacuties for binocular disparity and relative motion are the focus of the present study because they entail such exquisite sensitivity to spatial relations. We examine the potentially relevant spatial relations in more detail in the *Spatial Primitives: Frames of Reference for Spatial Position* section.

Statistical Uncertainty of Spatial Signals

Proposition 8: *Visual acuity is limited by the statistical uncertainty of spatial signals.*

All physical imaging systems are inherently subject to noise in the stochastic characteristics of photon capture, and this physical variability limits the precision of spatial position measures regardless of the spatial density of photodetectors and the optical point-spread function (den Dekker & van den Bos, 1997; Nowakowski

& Elbaum, 1983). Such physical and statistical limitations also apply to visual acuity. In addition to the physical noise inherent in retinal phototransduction, visual spatial discriminations are also limited by the spatiotemporal contrast modulation transfer functions of the eye and nervous system—dependent on the spatial density of receptors, the point-spread function of the eye, differential luminance resolution by photoreceptors and neurons, and reliability of signal transmission over neurons and synapses (see Geisler, 1989). The resulting net limits on spatial discriminations can be evaluated by psychophysical discriminations, providing a composite measure of the statistical discriminability of spatial positions.

Hyperacuity is a powerful tool for studying spatial vision because of the strength of the conclusions it permits about the mechanisms of spatial vision. First, as Westheimer (e.g., 1979b) and others have pointed out, hyperacuity is a sensitive performance index of visual spatial mechanisms. Second, information about spatial position must be represented in the relative stimulation of neighboring photoreceptors. Hyperacuties involve discriminations of positional differences significantly smaller than the separation between photoreceptors—spatial information that cannot be encoded simply as a function of which photoreceptors are stimulated and which are not. Third, the spatial information contained in the distribution of optical stimulation must be maintained with little loss of precision, little increase in the random noise affecting the position signals. The signal-to-noise ratio of the initial retinal information cannot be lost and then later restored by subsequent neural operations. Hyperacuity is, therefore, an especially impressive achievement in view of the inevitably increasing entropy of successive neural representations.

As a concrete illustration, consider the problem of discriminating two small spots of light that differ only in their spatial positions as illustrated in Figure 5. For the illustration in Figure 5, made from a computer video monitor, the receiver consists of a rectilinear array of spatially contiguous square-shaped detectors. Suppose that a single spot of light is presented on each of a series of trials at one of these two alternative spatial positions and that an observer's task is to decide on each trial between these two alternative positions. The positions of these stimuli are discriminable on the basis of the spatial distribution of intensities. Without noise, the positions theoretically could be estimated with arbitrary precision. Spatial discriminations are made more difficult, however, by the inherent random noise in the intensity measures by the individual photodetectors (as represented in Figure 5 by added random noise in the gray-scale value at each pixel). Thus, this spatial discrimination is necessarily a statistical problem demanding an efficient use of information contained in the spatial distribution of measured intensities.

Three classes of methods may be used to estimate the spatial positions of such signals in noise: (a) predictive filtering, (b) averaging and smoothing, and (c) correlations between multiple redundant images. Each of these uses a different form of redundant information in the spatiotemporal intensity distributions. Each method is usable by a theoretically ideal observer sensitive to all available image information, and each is potentially usable by biological systems. The spatial precision of all these methods is limited, however, by the statistical information in the images. Visual acuity is also limited by the image statistics, so it is instructive to examine this information.

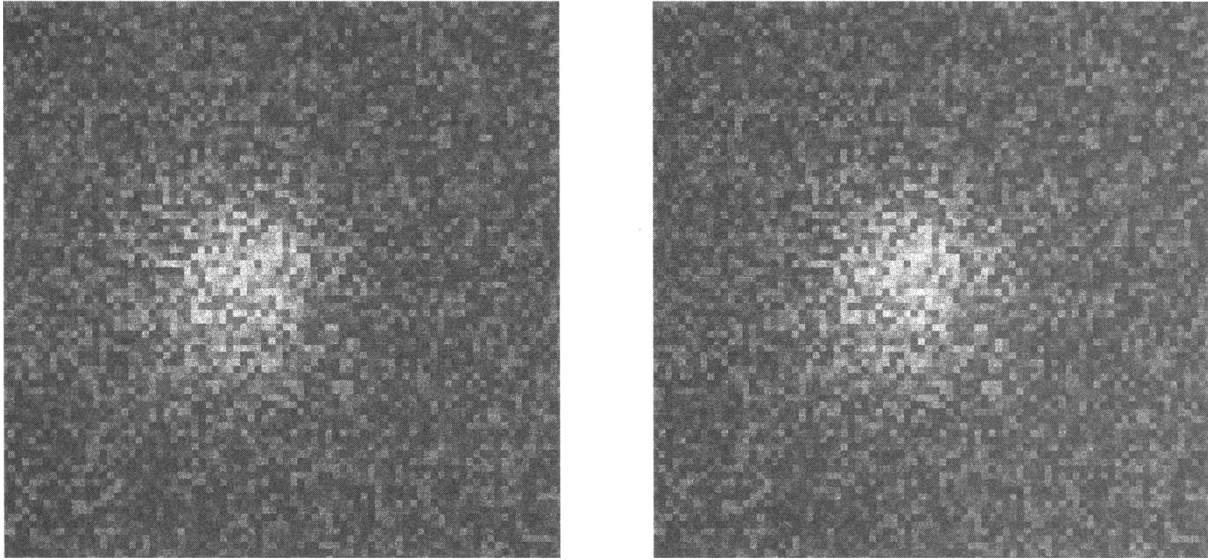


Figure 5. An illustrative example of spatial discrimination. A two-dimensional Gaussian blob of intensity has been superimposed on an array of randomly varied gray-scale values. The blob in the right panel is shifted three pixels to the right relative to its position in the left panel.

Predictive filtering methods involve comparisons (e.g., cross-correlations) of the spatial pattern of intensities on each trial with the two alternative signals. Such filtering is applicable when the signal parameters—for example, spatial positions, sizes of the spots, spatial intensity distributions—are known beforehand or when these parameters remain constant over a series of trials, permitting an adaptive learning process to tune selective filters. For such cases with known signal parameters, a linear discriminant function may be constructed for the statistical problem of deciding between the two alternative hypotheses. Such a filter provides an ideal observer with all the available statistical information. Many vision problems, however, do not permit use of such filters because the spatial positions and other characteristics of the signals vary widely from one occasion to the next and cannot be anticipated.

Alternatively, averaging and smoothing methods are based on systematic departures from the background noise distribution. For the simple case illustrated in Figure 5, the approximate location of the signal can be estimated by the higher than average image intensities in a region and by the intensity gradient over neighboring detectors. For example, the position could be estimated by the centroid of the intensity distribution as in a statistical frequency distribution. The position can also be estimated by fitting a smooth curve to the obtained intensity distribution as an estimate of the true signal envelope. Averaging and smoothing can be effective in removing signals from independent additive noise because the mean values of the signal and background noise tend over space and time toward reliably different values. Position discriminations improve in the same way that the statistical confidence interval of an estimated mean decreases inversely with the square root of the sample size.

A third class of spatial measuring methods involves correlating multiple coherent images of the same objects from different vantage points—as in holography and synthetic aperture imaging

systems. The 3-D structures of mutually camouflaging objects may be revealed and distinguished when a second image of the same scene is formed by moving the receiver relative to the objects. Cross-correlations between successive or stereoscopic images depend on the relative 3-D positions of the object features and on the motions that transform spatial positions and relations between images. Such cross-correlations may achieve exquisite spatial resolution, deriving from the coherence of multiple images of the same objects. In vision, such coherent structure occurs in images of 3-D scenes and is revealed by stereopsis and 3D motions of observers and objects. Ample experimental evidence demonstrates that human vision is very sensitive to the coherence of stereoscopic (e.g., Julesz, 1971; Tyler & Julesz, 1978) and moving images (e.g., Doner, Lappin, & Perfetto, 1984; Lappin & Bell, 1976; van Doorn, Koenderink, & van de Grind, 1984).

Cross-correlations between two discrete images, $I(x, y)$ and $I'(x, y)$, may be computed by the following formula:

$$C(\tau) = \sum_{-X}^X \sum_{-Y}^Y I(x, y) \times \tau[I'(x, y)], \quad (4)$$

where $\tau[I']$ is a transformation that maps image I' onto I , and where the spatial region of the cross-correlation is given by $x \in [-X, X]$ and $y \in [-Y, Y]$. The formula for cross- and auto-correlations usually is given for image translations, specified by horizontal and vertical displacements added to the x and y image coordinates, but correlations may be defined as well by other transformations that map one image onto another. The group of motions in 3-D space is of particular interest.

The fact that the cross-correlations are defined on the parameter space, τ , of the transformations rather than the initial image coordinates, x and y , is relevant to the descriptions of the spatial structure of images. First, the transformation parameters may represent motions in 3-D space rather than the 2-D image space.

The image displacements of multiple object features over successive frames define a vector field with a 3-D structure jointly determined by the shape of the object and its motion relative to the eye. Koenderink and van Doorn (1991, 1997), Tomasi and Kanade (1992), and Morita and Kanade (1997) have shown how such multi-image data may be represented very simply as bilinear functions of the object shape and its motion, both of which are sets of vectors in 3-D space. Thus, both the shape parameters and the motion parameters may be computed, up to an affine scaling ambiguity, by standard linear methods (e.g., singular value decomposition and principal-components analysis) from two or more images of sufficiently many object features.

A second interesting aspect of this description of space is that the resolution of spatial structure and motion may be dramatically improved over the spacing of photoreceptors or image pixels when the two images are highly correlated. In holography and other similar physical systems, the coherent image structure is provided by a high-frequency carrier, and the spatial precision derives from temporal phase relations. In the present examples, however, the coherence derives partly from the smooth surfaces on which the texture features are located and perhaps as well from coherent spatial relations among separate photoreceptors. Such long-range spatial coherence in the retina has not been well-known, but the present findings of hyperacuity for binocular disparity and relative motion point to such spatial coherence of retinal mechanisms (see also Lappin & Craft, 1997).

Whether this spatial structure is estimated by cross-correlation or some other method, the precision is statistically limited by the relative magnitudes of signal and noise. Redundant information is provided by additional image frames and by additional texture features of the same object, and this redundancy improves the spatial resolution. Multiple coherent images permit position discriminations much closer to the lower limits imposed by incoherent random noise than is otherwise possible with single stationary images.

All of the preceding methods of spatial localization operate by detecting the coherent structure of images, and they are limited by the statistical coherence of the images, between multiple images as well as within a single image. Such methods can dramatically improve signal-to-noise ratios and spatial resolution, but neural processing of multiple images cannot reduce spatial uncertainty below the limits imposed by the entropy of the initial retinal images. The following proposition applies this idea to inferences about the mechanisms underlying spatial acuity:

Proposition 9: *Spatial uncertainty increases with successive neural representations. A corollary is that hyperacuity represents an upper bound on the spatial uncertainty of the retinal stimulation.*

This proposition is a methodological relative of Brindley's (1970) rationale for "Class A" experiments: Behavioral discrimination implies a difference in both retinal and other neurophysiological responses. Similarly, acuity entails discriminations of differences in spatial position and is limited by the statistical discriminability of signals from different retinal positions. Neural operations may be required to reveal the statistical and spatial structure implicit in the binocular and moving images, but these neural processes cannot increase the statistical discriminability of spatial positions beyond the limits given by the binocular and moving images. This property is a physiological analog of the

second law of thermodynamics, that the entropy of a closed system cannot decrease. Hyperacuity requires maintenance of low spatial uncertainty throughout all stages of neural representation and processing. Spatial signals that are statistically indistinguishable at the retinas cannot be made to differ at a later neural representation, nor can their statistical discriminability be increased.

Spatial Primitives: Frames of Reference for Spatial Position

Spatial positions are defined in relation to some reference frame. The reference frame pertinent to a particular acuity task may be identified by the group of image transformations under which discriminations of spatial position remain invariant. That is, if thresholds for discriminating a difference in the position of a given target feature increase with the addition of certain random image transformations, then these transformations must have perturbed the spatial position of the feature relative to a relevant frame of reference, but if acuity remains unaffected by a group of image transformations, then the visually defined position of the feature must be unchanged by these transformations. Thus, the following proposition can be made:

Proposition 10: *The primitive spatial structure that serves as the visual reference frame for the spatial position of a given image feature may be experimentally distinguished by the group of image transformations under which acuity for relative position remains invariant. This visual reference frame is called a spatial primitive.*

The experimental method indicated in this proposition was used in a recent study by Lappin and Craft (1997). Uncorrelated dichoptic random image transformations were used to identify the monocular spatial primitive underlying acuity for binocular disparity.

The present study used hyperacuity for binocular disparity and relative motion to identify spatial primitives for visual position. The experiments were designed to distinguish among a set of alternative hypotheses specified by the following proposition:

Proposition 11: *Spatial primitives may be classified by the order and dimensionality of space differentials of the local image structure.*

This proposition is based on Koenderink and van Doorn's (1992b) scheme for describing and classifying measures of local image structure in terms of the order and dimensionality of spatial derivatives, as described above in the Retinal Images of Environmental Objects section. This description of the differential geometry of the image is useful both because of its relationship with the differential geometry of the object surface and also because spatial relations of various orders of complexity differ from one another in terms of their invariance under spatial transformations of the image.

Five alternative hypotheses about the visual spatial primitives correspond to the zeroth, first, and second orders of spatial derivatives in one and two dimensions. (Zero-order relations are dimensionless, and two orders of differentiation in either one dimension or two dimensions yield four additional combinations.) The literature on spatial vision contains examples of theoretical ideas corresponding to all five of these alternatives. Lappin and Craft (1997) recently tested these hypotheses in stereoacuity experi-

ments, and the present study extends this investigation to motion detection. These five hypotheses are the following, where the number specifies the order of differential complexity.

H0: *Spatial positions of image features are visually specified by retinal coordinates, extrinsic to the optical pattern, based on zero-order spatial structure. Visual acuity for differences in spatial position is reduced by random image translations.*

H1: *Spatial positions of image features are visually specified by first-order (pairwise) spatial relations intrinsic to the optical pattern. Visual acuity is robust under random rigid translations that disrupt zero-order image structure.*

H2.1: *Spatial positions of image features are visually specified by second-order intrinsic image structure involving first-order space differentials jointly in two dimensions. Visual acuity is robust under random image expansions (divergence) as well as translations and 2-D rotations (curl), which disrupt zero- and first-order spatial relations but leave higher order structure invariant.*

The spatial relations specified by Hypothesis H2.1 involve just first-order space differentials in any given direction, but they constitute a form of second-order structure because they involve simultaneous, interdependent relations among three neighboring features or regions in a triangular arrangement. These structural relations are more complex than those in Hypothesis H1; they involve relations among three neighboring image points rather than pairwise relations pertinent to H1. The second-order structure in H2.1 remains invariant under isotropic scale changes that alter the first-order relations pertinent to H1.

H2.2: *Spatial positions of image features are visually specified by second-order relations involving second-order space differentials in any given direction within the optical pattern. Visual acuity is robust under random translations, 2-D rotations (curl), and image expansions (divergence), and discriminations between planar and curved surfaces are invariant under 3-D rotations that alter the slant in depth.*

The complexity of the spatial primitives is the same for both Hypotheses H2.1 and H2.2, both involving relations among three image points or regions. In H2.2, these three points extend in a single direction and entail the symmetry, relative separations, or relative intensities on either side of a given central reference feature. Two different forms of such second-order relations involve differential compression, involving changes in the relative positions or symmetry among three collinear points, produced by image motions in the direction of the points' alignment, and quasi-collinearity, which involves changes in the degree of collinearity among three approximately collinear points, produced by image motions orthogonal to the direction of alignment. Both of these image changes are produced by 3-D rotations of curved surfaces.

H4: *Spatial positions of image features are visually specified by fourth-order relations involving second-order space differentials jointly in two directions within the 2-D optical pattern. Visual acuity for the relative position of a surface feature remains robust under 3-D rotations that alter the position and orientation of the surface in three dimensions.*

The spatial relations pertinent to Hypothesis 4 involve relations among at least five neighboring image regions, with a point on

either side in two dimensions around a given central reference point. These are second-order space differentials in any given direction, but they are measured in relation to those in a second dimension. These "fourth-order" relations may be experimentally distinguished from the second-order relations pertinent to Hypothesis H2.2 because they permit acuity for arbitrary surface shapes invariant under changes in 3-D orientation. The two forms of second-order structure pertinent to Hypothesis H2.2 distinguish surfaces that are planar versus curved in any given direction, but these second-order relations are not sufficient for discriminating surface shapes, which involve relative curvatures in two dimensions simultaneously. These fourth-order spatial relations as well as the two-component second-order relations are illustrated in Figure 3.

A schematic illustration of the spatial support and complexity involved in all five of these hypotheses is given in Figure 6. The lower part of this illustration shows possible local operators sensitive to each of these five spatial primitives, formed from spatial derivatives of a Gaussian kernel. These local operators are shown here only as a concrete illustration of how higher order spatial relations can be detected by simple local linear operators. The potential role of such operators was not tested in the present experiments.

An implicit idea in Proposition 11 is that visually defined spatial positions may depend on spatial derivatives of the surrounding luminance distribution. In the present experiments, we did not directly test such potential luminance effects. The relative positions of neighboring lines and points were manipulated, but not their luminous contrasts. If visual acuity is influenced by changes in relative positions, then we might expect that it would also be affected by changes in their relative luminances, and if acuity is independent of certain changes in the relative positions of surrounding features, then we might also expect that it would be unaffected by corresponding changes in their luminances. These assumptions warrant tests in future experiments.

The visually relevant spatial neighborhood of a given image feature may be rather large. The present experiments indicate that this neighborhood is at least 4° of arc radius, and other experiments indicate that it may be as large as 12° (Lappin & Ahlström, 1996). Such extensive spatial relations might be thought to involve independent local mechanisms, but the present results show that the visual mechanisms do not provide locally independent spatial signals. Rather, visual information about the position of a given feature is referenced to its neighbors. The variability of these spatial signals is a roughly constant fraction of about 0.25% of the distance between features, over a range of 0.5 to 4° of arc in the present study.

The differential order of the local image structure is not assumed to imply anything specific about the underlying visual mechanisms. The aim is to characterize the spatial structure of the visual input. We postpone discussion of visual mechanisms for detecting this input.

One potential misconception of the visual mechanisms might be that measures of higher order spatial derivatives require computation of differences between lower order measures. Koenderink (1990, pp. 37–40) and Koenderink and van Doorn (1992b) have shown how higher order differentials can be computed directly by linear integration without first computing the lower order property. The fact that higher order derivatives do not require measures of

Five Structures

		Derivatives		
		0	1st	2nd
Dimensions	0			
	1			
	2			

Five Operators

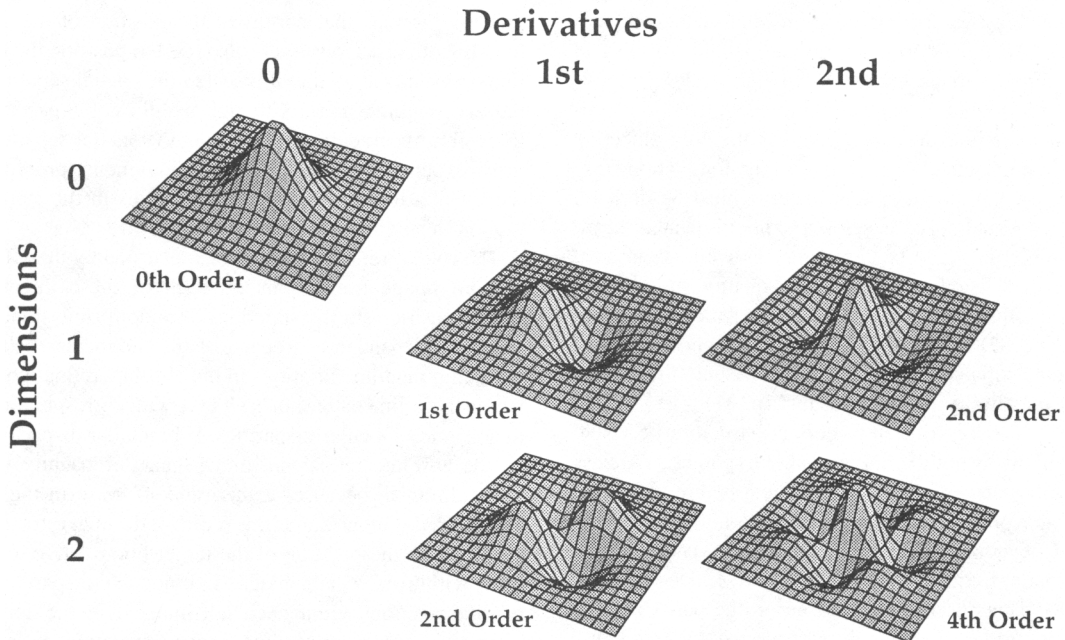


Figure 6. Schematic illustration of the spatial support and complexity involved in five alternative hypotheses about the spatial primitives of the visual input (top). The lower part of this illustration shows possible local operators sensitive to each of these five spatial primitives, formed from spatial derivatives of a Gaussian kernel. The potential structure and contribution of such local operators were not tested in the present experiments. They are illustrated here only as a concrete illustration of how higher order spatial relations can be detected by simple local linear operators (see Koenderink & van Doorn, 1992b). Reprinted from *Vision Research*, 37, J. S. Lappin and W. D. Craft, "Definition and Detection of Binocular Disparity," p. 2954, Copyright 1997, with permission from Elsevier Science.

lower order properties is critically important in the present experimental methods, where we identify the differential order of a primitive spatial relation by its invariance under random perturbations of lower order properties. The difference between two or more independent lower order measures will necessarily have higher variance than the lower order measures. If visual acuity for a higher order image property remains robust under added noise in lower order image properties, then this visual acuity cannot have been derived by comparing the lower order measures.

Next, we use this methodological rationale to test the preceding hypotheses about the spatial primitives of visual input.

Experiment 1: Extrinsic Versus Intrinsic Reference Frames

The first question is whether the visual reference frame for spatial position is extrinsic to the optical patterns—specified by the anatomical mosaic of retinal photoreceptors, independent of the optical stimulation—or intrinsic—specified in relation to surrounding image features, independent of retinal position per se. This distinction is equivalent to that between zero-order and higher order image structure.

A zero-order description of the image is a scalar field, giving the luminous intensity of the image at each spatial position. These spatial positions are necessarily extrinsically defined, independent of the particular intensities at these positions and of their pattern over space. Visual spatial information about spatial position and spatial relations among image features is thought to be anatomically defined by the retinal mosaic of photoreceptors. This zero-order description sometimes has been taken as the only objective physical description of optical stimulation, as the foundation from which visual representations of spatial patterns must be derived. This conception of visual space is seldom explicitly stated, but it is implicitly assumed in many contemporary as well as classical discussions of spatial vision—involving hyperacuity, motion perception, stereopsis, illusions, shape and size constancy, and depth perception. Marr (1982) is one of the few authors who have been explicit about this spatial reference frame for visual stimulation, and he said that its validity is “in no doubt!” (p. 31).

However obvious the extrinsically defined structure of visual space might seem, this is not the only possible hypothesis. A class of alternative hypotheses proposes that spatial structure is defined intrinsically by the optical stimulation, by relations among the image features. One example is the hypothesis that visual input is encoded by its spatial frequency spectrum (see R. L. De Valois & De Valois, 1988). This description is based on the periodicity of pairwise relations among image points separated by fixed distances and directions, and these pairwise relations are invariant under image translations.⁹ Additional hypotheses about the intrinsic spatial structure of visual input are examined more thoroughly later.

An important question raised by the hypothesis that visual space is intrinsically defined concerns the visual mechanisms for measuring image structure. The idea that spatial information is independent of retinal position might seem to ignore the basic problem of visual encoding of spatial relations as well as the critical role of retinal anatomy and physiology. The extrinsic and intrinsic hypotheses do not differ on whether the retinal photoreceptors are relevant to spatial vision, but they do differ on the degree to which

visual space is structured by retinal coordinates. Intrinsic image structure as described by partial derivatives of the form $\partial r_i / \partial s_j$ is invariant under translation; absolute positions need not be specified. As an analogy, consider specifying a circle at a given point (h, k) —as a set of points satisfying a relation $(h - x)^2 + (k - y)^2 = r^2$, or as a set of points equidistant from the given point. The former entails extrinsic Cartesian coordinates, but the latter can be constructed intrinsically with a compass.

Experiment 1 was designed to test the following two alternative sets of hypotheses:

H0: *Spatial positions are visually specified by retinal coordinates, extrinsic to the optical pattern. Visual acuities are reduced by random image translations.*

H1–H4: *Spatial positions are visually specified by intrinsic image structure. Visual acuities are robust under random image translations.*

These two hypotheses were tested primarily by the hyperacuities for detecting positional differences defined by binocular disparity and by motion. In the stereo task, observers adjusted the position of a vertical line so that it appeared to be in the same fronto-parallel depth plane as two other flanking lines—attempting to eliminate binocular disparity of the central target line. In the motion detection task, observers tried to eliminate small random jittering motions that perturbed the position of the central target line. In both tasks, observers also tried to position the central target line on each trial so that it bisected the spatial separation between the two flanking lines. Although we did not expect the observers to exhibit hyperacuity for bisecting the spatial separations used in these experiments, these bisection judgments provide a standard with which to compare the hyperacuities in the stereoacuity and motion detection tasks.

The role of retinal coordinates was tested by the effects of noisy random image translations on each of these three acuities for spatial position. In the stereo task, random rigid translations were applied independently to each of the monocular half-images, introducing random variations in the absolute retinal positions of the whole three-line pattern in each eye, with corresponding variability in absolute binocular disparities. If binocular disparity were based on the absolute retinal positions, then stereoacuity would necessarily diminish because of this random noise in the retinal positions. On the other hand, these dichoptic image translations have no effect on the position of the target line relative to the flanking lines within each half-image. If binocular disparity is based on relative positions within each half-image, then the stereoacuity will be unaffected by the random image translations.

⁹ A popular contemporary approach that combines sensitivity to both spatial frequency and extrinsically defined local spatial position uses Gabor functions, which are formed by multiplying sine and cosine functions by 2-D Gaussians. Such local operators are sensitive to aspects of both intrinsic spatial frequencies and to either extrinsic spatial positions or intrinsic phase structure. Such local operators might encode spatial positions either by the intrinsic image structure, relative to the outputs of overlapping and neighboring operators, or by reference to the extrinsic reference frame of the retinal mosaic. This extrinsic conception seems to be implicitly assumed in most of the literature, but this assumption is not required in using these operators.

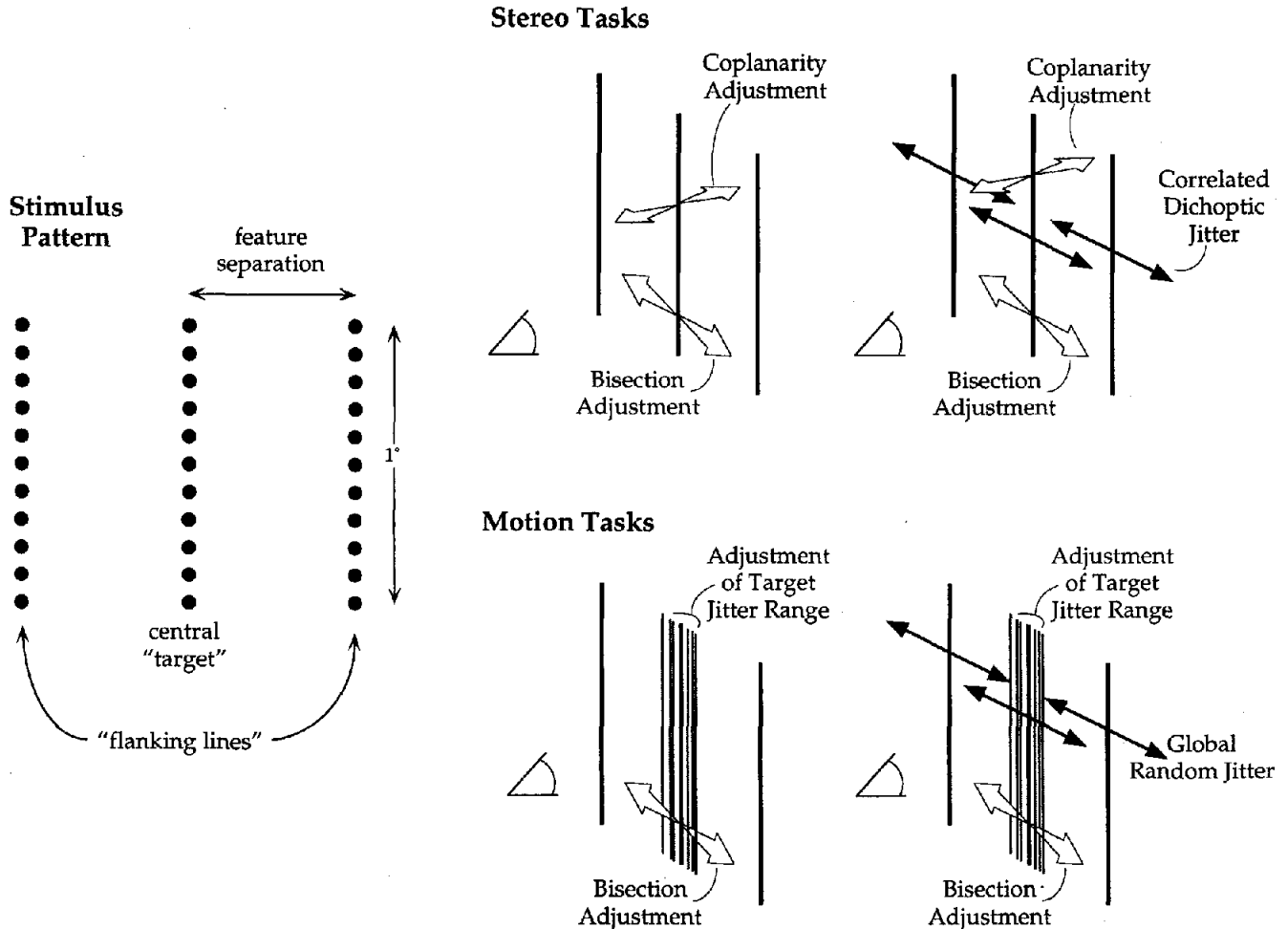


Figure 7. Schematics of the stimulus pattern, tasks, and manipulations in Experiment 1. In the stereo conditions, observers adjusted the central target line to coplanarity with the flanking reference lines and also simultaneously adjusted the target line to bisect the distance between the flanking lines. In the motion conditions, observers eliminated relative motion of the target line with respect to the flanking lines and simultaneously adjusted the target to bisect the distance between the flanking lines. In two different display conditions with both of the stereo and motion conditions, the global pattern of central target plus flanking lines was either stationary or jittered. The stationary condition is illustrated in the two middle panels, and the jittered conditions are illustrated in the two right panels. The observers' adjustments of the central target position were added to changing positions of the global pattern.

A corresponding rationale applies to acuity for differential motion of the target line relative to the flanking lines. Random globally rigid motions were added to the whole three-line pattern, and the observer tried to detect and eliminate uncorrelated motion of the target line. If acuity for detecting differential motion of the target line remains the same either with or without this added global motion, then this differential motion must be referenced to the surrounding optical pattern.

Method

Stimulus patterns and apparatus. A schematic representation of the stimulus patterns appears in Figure 7. Stimulus patterns consisted of three vertical lines in a horizontal array, each line being 1° in length and made

up of 11 equally spaced, easily distinguished points, with the lines being separated from each other by 0.5° , 1° , 2° , 3° , or 4° of visual angle.

The patterns appeared as bright white dots on a dark background and were viewed through a haploscopic arrangement of two Tektronix (Beaverton, OR) 608 display oscilloscope monitors (P34 phosphor) controlled by a laboratory computer and a D/A converter with 16-bit resolution both horizontally and vertically. At a viewing distance of 57.3 cm, the effective resolution was 0.44 arcsec. The computer program allowed independent manipulation of each eye's image and of each feature within an image, control of frame and trial length, and the collection and storage of response data.

Tasks and conditions. Two adjustment responses—bisection and detection—were made on each trial as illustrated in Figure 7. In the stereo bisection task, observers used a joystick to adjust the horizontal position of the central "target" line to bisect the distance between the two flanking

lines, and in the stereo detection task they used another joystick to adjust the relative depth of the target to make it coplanar with the flanking lines in the fronto-parallel plane. In the motion task, bisection adjustments were essentially the same as in the stereo task. In the motion detection task, observers used the second joystick to eliminate horizontal jitter added to the target relative to the two flanking lines (which were rigidly connected to one another).

In two different display conditions, the two flanking reference lines were either stationary or jittered, with rigid image jitter added to the horizontal and vertical positions of both the target and frame. This image jitter was created by randomly repositioning the entire pattern every 100 ms within an 11×11 grid of equally spaced and equally probable positions, 12×12 arcmin in width and height. The root-mean-square (rms) frame-to-frame displacement was 5.37 arcmin both horizontally and vertically—7.59 arcmin total. For the motion task, this image jitter was binocularly correlated; for the stereo task, the image jitter was dichoptic, corresponding to a random jitter in three dimensions. This image occurred in a repeating cycle of 50 frames, but the repetitions went unnoticed.

The range of adjustment controlled by the joysticks was ± 9.3 arcmin in the horizontal direction in all conditions. In the motion detection task, the joystick controlled the range of a uniform distribution of possible positions sampled randomly by the target every 100 ms. In the stereo detection task, the second joystick adjusted the depth of the target line within a range of ± 9.3 arcmin disparity relative to the two flanking lines. The initial position of the target, or its initial relative motion, was varied by small amounts from trial to trial, so that observers could not reduce the adjustment variability simply by remembering the joystick position from trial to trial. Five values of feature separation were investigated in both the stereo and motion tasks, with 0.5° , 1° , 2° , 3° , or 4° of arc between the target and flanking lines.

Procedure. The stereo and motion tasks were combined with the stationary and jittered display conditions to create four different task and jitter-condition combinations, which were interleaved in counterbalanced order in successive sets of four sessions. Each experimental session consisted of five blocks of trials dedicated to a single task-jitter combination (e.g., all blocks might be jittered stereo conditions), with each block dedicated to a different feature separation. Each block consisted of 5 trials. Each of the 5 trials produced values for both bisection and detection judgments. A block of trials typically took 5–10 min, with 2–3 min of rest between blocks. Experimental sessions were typically conducted once or twice a day, and four sessions were completed for each condition, producing 20 trials at each of the 20 combinations of task, jitter, and feature separation.

Observers. Three well-practiced observers, including Warren D. Craft (WDC), served in all experiments reported in this article. All had normal or corrected-to-normal acuity.

Results and Discussion

Acuity was measured by the standard deviation of the 20 positional adjustments in each condition. Figure 8 shows these acuities as a function of feature separation in each condition for each of the three observers. Standard errors of these estimated standard deviations were computed by a bootstrap method (Efron & Tibshirani, 1991).

These acuities can also be described by Weber fractions, dividing the standard deviation by the spatial separation between adjacent features. Figure 9 presents the same acuities as in Figure 8 as Weber fractions. Whereas Figure 8 indicates that changes in feature separation have a much greater effect on the bisection acuities than on the detection acuities for eliminating relative motion and relative disparity, Figure 9 reveals a simpler result: Increasing feature separation had similar effects on the Weber fractions for

the bisection as well as both the stereo and motion detection judgments. The Weber fractions were relatively stable over feature separations between about 1° and 4° and were slightly worse at smaller separations. For the separations between 1° and 4° , the average Weber fractions for the bisection acuities were 0.78% and 0.83% for the stationary and jittered patterns, respectively. The stereo and motion detection hyperacuities for changes in these target positions, however, yielded much smaller Weber fractions—0.22% and 0.29% for the stationary and jittered stereo patterns, respectively, and 0.06% and 0.18% for the corresponding motion patterns. In all these conditions, acuities depended on the distance between features.

The principal result is that hyperacuities were robust under large amounts of image jitter. Overall, acuities were significantly better when the reference frame was stationary than when it was jittered, but these effects on acuity were very small in relation to the random positional variations produced by the image jitter. With 1° feature separation, for example, random image noise of more than 340 arcsec in the horizontal positions produced an increase in the rms average stereo threshold from 12.4 to 13.2 arcsec and in the motion detection threshold from 3.1 to 10.1 arcsec. Spatial positions evidently are visually defined mainly by intrinsic optical structure rather than extrinsic retinal coordinates.

This conclusion differs from that of Westheimer and McKee (1979) and McKee et al. (1990). They and other authors have suggested that spatial position is an essentially local property, encoded by visual mechanisms tagged with local signs of retinal position. McKee et al. acknowledged that optimal acuities for binocular disparity and relative motion require a spatial reference feature, and they suggested that such cues may serve to eliminate positional uncertainty caused by eye movements. The results of Experiment 1, however, indicate that these hyperacuities must be based on spatial relations among neighboring features rather than absolute retinal positions: (a) Hyperacuity was maintained under random image perturbations much larger than those of microsaccadic eye movements and much larger than the stereo and motion hyperacuities, and (b) both stereo and motion detection hyperacuities as well as 2-D bisection acuities were approximately proportional to the feature separation.

The qualitative and quantitative similarities among these various acuities provide clues about the visual reference frame for spatial position and about the processes by which hyperacuity is achieved. Notice first that performance was quantitatively similar for both stereo and motion conditions. Second, both the stereo and motion detection acuities were considerably better than the bisection acuities, although these 2-D relative positions must serve in some sense as the input for detecting stereoscopic disparity and relative motion.

The difference between the bisection acuity and the stereo and motion detection hyperacuities for the same patterns is similar to the difference found by Westheimer and McKee (1979) and by McKee et al. (1990). Because the bisection and detection tasks exhibit quantitatively different dependence on the spatial separation between features, different visual mechanisms were thought to be involved in these two forms of acuity, with both “founded on an excellent capability to localize the image of a feature on the retina” (Westheimer & McKee, 1979, p. 620). When the data are recast in terms of Weber fractions as shown in Figure 9, one sees that both the bisection acuities and detection hyperacuities share the same

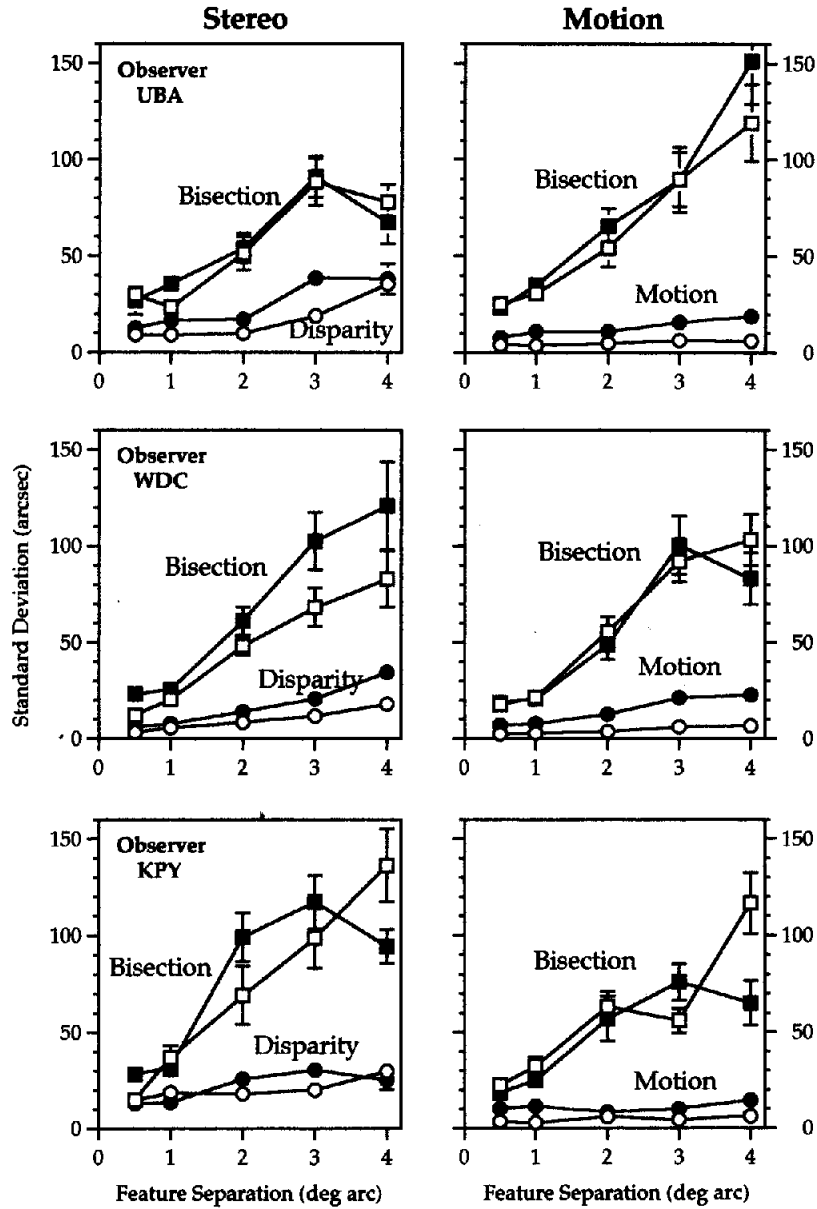


Figure 8. Results from Experiment 1, plotting positional variability (standard deviation in arcsec) versus feature separation for three observers under both stereo (left panels) and motion conditions (right panels). Filled symbols designate two-dimensional jitter conditions; open symbols designate stationary-frame conditions. Error bars designate $\pm 1 SE$ estimated with a bootstrap method; missing error bars were simply too small to appear.

functional dependence on feature separation. A basic theoretical question, therefore, concerns the mechanisms that produce the superiority of the stereo and motion hyperacuties. This issue is revisited below.

Experiment 2: Spatial Position Under Image Expansions

Experiment 2 examined the effects on acuity of another group of image transformations—random expansions and contractions—of one monocular half-image relative to the other in the stereo patterns and of successive images in the motion patterns. The effects

of these image transformations are relevant to the following questions about acuity.

First, despite the results of Experiment 1, might retinal coordinates still provide the fundamental spatial reference frame and might the visual system simply be very rapid and virtually noise-free in comparing the relative retinal positions of neighboring features? We can also test H_0 by reducing the effectiveness of information about simple differences in retinal position.

Second, does hyperacuity for detecting binocular disparity depend on the special case of a fronto-parallel plane? Does a neigh-

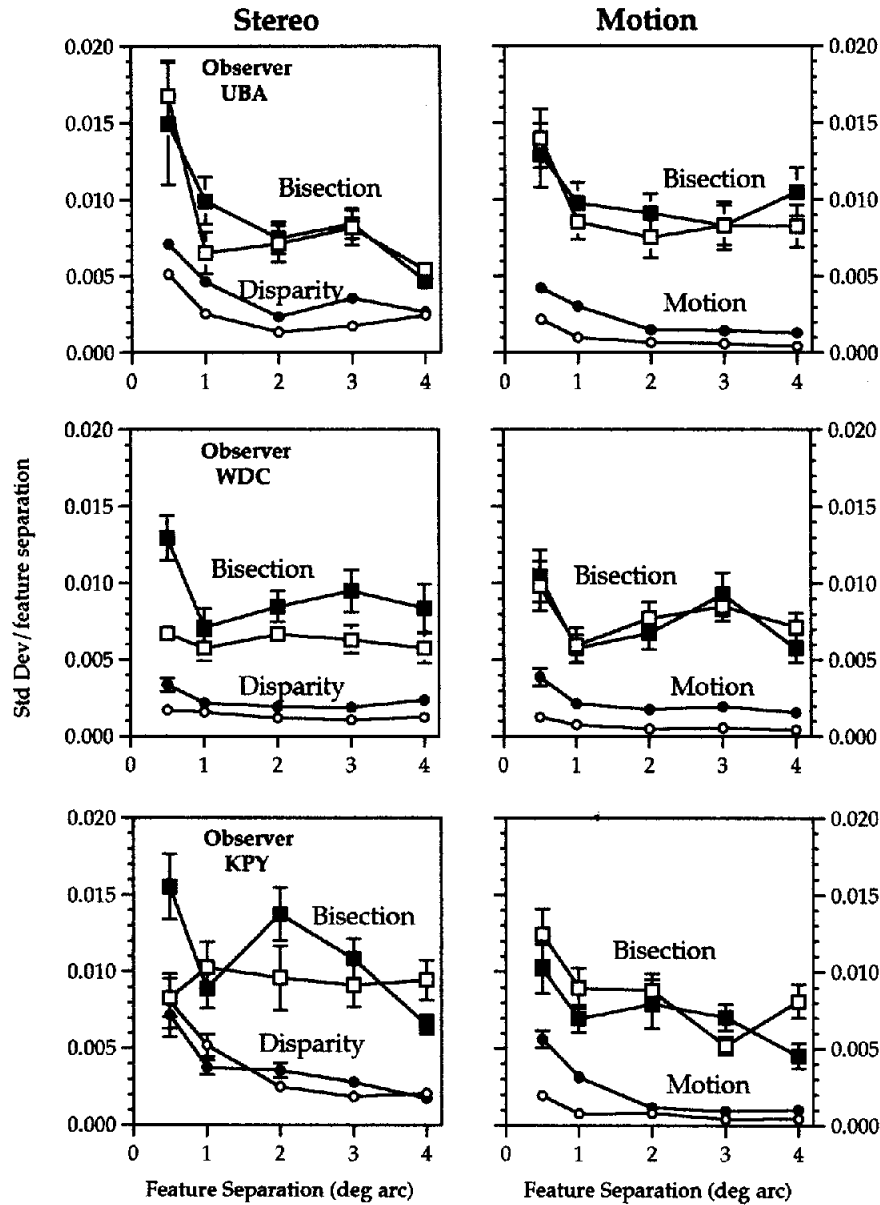


Figure 9. Results from Experiment 1, plotting Weber fractions (standard deviation [Std Dev] divided by feature separation) versus feature separation for three observers under both stereo (left panels) and motion conditions (right panels). Filled symbols designate results for the globally jittered patterns; open symbols designate results for the stationary conditions. Error bars designate ± 1 SE estimated with a bootstrap method; missing error bars were simply too small to appear.

boring image feature simply define a reference for the relative disparity of the target feature? If a surface is slanted in depth from the fronto-parallel image plane, then stereoscopic discriminations between planar and curved surfaces cannot rely on the disparity between the monocular separations of a pair of features. Similarly for detecting relative motion, if a curved object rotates in depth, expanding or contracting its retinal images, then perception of surface curvature from motion could not be based only on changes in the first-order separations between pairs of features. Perhaps spatial positions are based on higher order relations among three or more neighboring features.

Experiment 2 offers a test of H1, that spatial positions are visually defined by first-order spatial relations, as compared with H2.1, H2.2, and H4, that spatial positions are visually defined by higher order spatial relations.

To test the potential visual role of second-order structure, we can disrupt first-order structure by random image expansions. The relative target position would now be defined by relations among three rather than two features. If the hyperacuitys in Experiment 1 were based on first-order relations, then the acuity thresholds would necessarily increase when the target position is defined by relations among three features and requires a comparison among

two first-order spatial relations. If these two first-order relations are visually independent, then the acuity thresholds should increase by a factor of $\sqrt{2}$ relative to those for the lower order components—because the variance of a difference between two independent variables equals the sum of the component variances. If this increase does not occur, then the lower order relations cannot be visually independent of one another.

Acuties and hyperacuties therefore provide a powerful method for identifying visual spatial primitives: Acuity thresholds for higher order spatial relations must be greater than those for lower order relations if the lower order information is visually primitive.

Method

Stimulus patterns. The stimulus patterns consisted of three parallel vertical lines, the same as in Experiment 1. Prior to the application of any image transformation, the separation between adjacent lines was 1° . In the motion conditions, horizontal image expansions of 0%, $\pm 1\%$, $\pm 2\%$, $\pm 4\%$, or $\pm 8\%$ (in separate conditions) were applied in a sequence of alternating expansions and contractions in successive 100-ms frames. In the stereo conditions, the right monocular half-image was expanded by 0%, 1%, 2%, 4%, or 8% (in separate conditions), while the left half-image was contracted by an equal and opposite amount. The resulting interimage disparities in the position of the target relative to the flanking lines were 0, 1.2, 2.4, 4.8, or 9.6 arcmin.

Tasks and conditions. The bisection and detection tasks were conceptually identical to those in Experiment 1, but with some perceptual differences. In the stereo conditions, the reference plane defined by the two flanking lines usually appeared to be slanted in depth. The stereo detection task required adjusting the target to be coplanar with the two reference lines, and the bisection task required bisecting the 3-D separation between the two flanking lines. The motion detection tasks were similar, involving patterns that expanded and contracted in the horizontal dimension as would have been produced by a plane rigidly rotating in depth, though in fact these patterns did not appear to move in depth. The motion detection task required the elimination of additional horizontal jitter uncorrelated with these expansions to maintain a constant relative position of the target between the two flanking lines, and the motion bisection task required centering the target between the two reference lines.

As in Experiment 1, rigid global horizontal and vertical translations were imposed on the target and reference lines in both task conditions. Because of the additional motion produced by the horizontal expansions, these random translations were approximately half the magnitude of those in Experiment 1, with an rms frame-to-frame displacement of 2.68 arcmin in both horizontal and vertical directions. In the stereo conditions, these random translations were binocularly uncorrelated (as in Experiment 1), and in the motion conditions, this random jitter was binocularly correlated but temporally uncorrelated. There were no stationary image conditions.

Procedure. The procedures were like those in Experiment 1. The magnitude of the image expansions was varied between successive blocks of trials within each session.

Results and Discussion

The results of Experiment 2 appear in Figures 10 and 11. Figure 10 gives the distributions of position adjustments for the stereo task (20 for each observer in each condition), shown to scale within the implied slanted plane as viewed from above. The standard deviations of these judgments are shown in Figure 11 for detection and bisection tasks in both stereo and motion conditions.

As may be seen, stereoacuity was robust over substantial binocular disparities in radial image size. The radial size disparities

were at least 10 times larger than the detectable disparities in relative target position. The largest radial size disparity, 9.6 arcmin, produced a reliable increase in the stereo detection thresholds, but the disparity thresholds still were more than an order of magnitude smaller than the global size disparity. The stereo detection thresholds were also much lower than the bisection thresholds, though the effect of the dichoptic image expansions on the bisection judgments differed among the three observers.

These stereoacuity results are consistent with Hypotheses H2.2 and H4 that visual positions are defined by second-order image properties. The maintenance of stereoacuity over these disparate expansions of the horizontal scale indicates that spatial positions are not defined by either zero- or first-order relations as postulated by Hypotheses H0, H1, and H2.1.

Motion detection thresholds, on the other hand, were more strongly influenced by these changes in the horizontal image scale. Surprisingly, the motion bisection judgments seemed to have been less influenced by these changing size differences than were the stereo bisections, apparently based on temporally integrated perceived patterns. More important, the superiority of the motion detection performance as compared with the bisection acuties disappeared for interimage differences of about 2.4 arcmin or more. For these motion tasks, the general conclusion must be that the perceived image positions and motions derive mainly from the first-order, or pairwise, spatial relations among features as specified by Hypothesis H1 or H2.1. Information about second-order structure seems to have been derived from these first-order relations. Nor did the random expansions and contractions of these patterns produce any subjectively perceived 3-D rotation or translation in depth, unlike the stereoscopic perceptions.

These discrepant results on the spatial primitives underlying binocular disparity and relative motion replicate the findings of Lappin and Craft (1997) on stereoacuity and of Lappin, Ahlström, Craft, and Tschantz (1995) on perceiving structure from motion. As in the present results, Lappin et al. found that detection of relative motion did not seem to be based on second-order changes in "differential compression," involving horizontally moving points. Second-order changes in "quasi-collinearity," however, with horizontal motions of vertically aligned points, did seem to constitute visually primitive spatial relations. A similar anisotropy, with greater sensitivity and perceived 3-D structure perpendicular as compared with parallel to the image motion, has also been observed in previous studies (e.g., Norman & Lappin, 1992; Rogers & Graham, 1983). Earlier work by Wallach and O'Connell (1953) and Johansson (1964/1994) also found that image deformations in a single dimension usually were perceived as nonrigid motions in the 2-D image plane rather than as rigid rotations in three dimensions.

Thus, the horizontal image expansions and simple spatial patterns examined in Experiment 2 may have been visually insufficient as information for perceiving 3-D structure and motion. The fact that these did not appear as rigid 3-D motions reinforces this possibility. Experiment 3 examined other spatial patterns and motions as support for perceiving higher order spatial relations.

Experiment 3: Spatial Position on Curved Surfaces

If local surface shape is specified by the 2-D structure of second-order derivatives of the interimage displacement field as

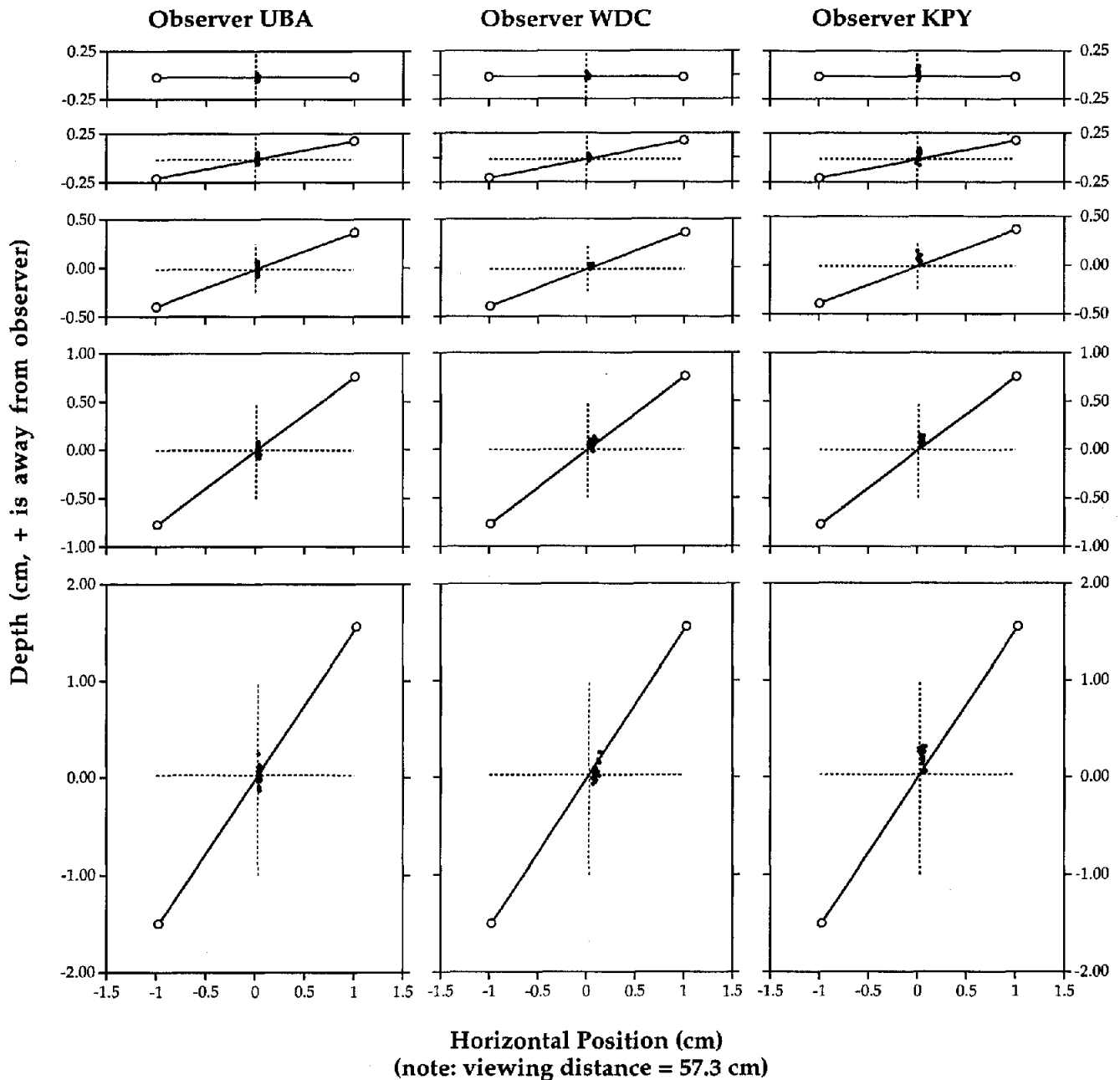


Figure 10. Three-dimensional positional settings for the stereo detection task in Experiment 2 for each observer and each interocular expansion condition. The solid line indicates the geometrically predicted implicit surface formed by the flanking reference lines due to the interocular image size difference. The dotted lines demarcate the actual center position for comparison. The interocular radial size difference for the five rows of data, from the top, were 0, 1.2, 2.4, 4.8, and 9.6 arcmin, respectively.

stated in Proposition 5, then vision may be directly sensitive to this image structure as a spatial primitive as postulated in Hypothesis H4.

For stereopsis, Experiments 1 and 2 rule out Hypotheses H0, H1, and H2.1 but do not distinguish between Hypotheses H2.2 and H4, both of which postulate sensitivity to second-order spatial derivatives of the binocular disparity field. The question distin-

guishing these two alternatives is whether visual measures of the disparity field are inherently 2-D, involving conjoint, nonindependent measures in two dimensions. Hypothesis H2.2 entails second-order differential structure, whereas H4 entails fourth-order structure. These two hypotheses differ about the perception of spatial positions on arbitrarily curved surfaces, because local surface shape is described by the relative magnitudes of second-order

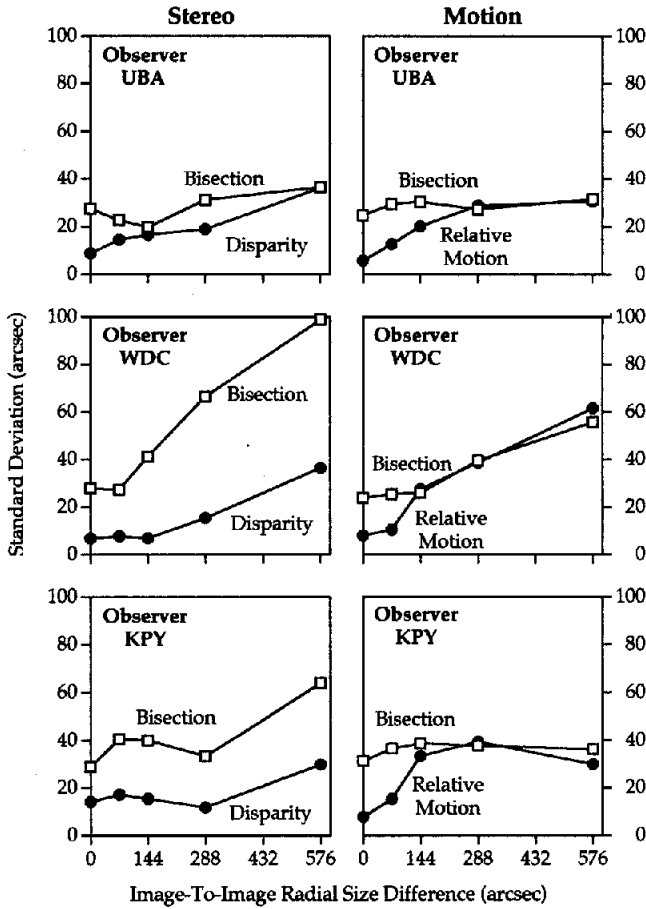


Figure 11. Results of Experiment 2, plotting positional variability versus image radial size difference. Motion detection acuities degenerated to the level of the corresponding bisection acuity at image size differences that left stereo detection acuities relatively unaffected.

derivatives in two dimensions. Hypothesis H2.2 predicts larger acuity thresholds for curved than for planar surfaces, whereas H4 predicts similar acuities for both surfaces.

For perceiving spatial structure from motion, Experiments 1 and 2 do not justify rejection of Hypothesis H1. If perception of spatial structure from motion were actually based on the fourth-order image properties, associated with smooth surfaces rotating in depth, then the essentially 1-D patterns presented in Experiments 1 and 2 may have been insufficient as images of such surfaces.

Experiment 3 compares acuities for positions on spherical and planar surfaces. When a spherical surface patch is slanted in depth, then the local shape and surface position are specified in the image by fourth-order differential structure and not by independent measures of second-order derivatives in two directions. Figure 12 illustrates the more complex spatial relations required to specify position on a spherical as compared with a planar surface.

Figure 12A shows a hexagonal array of points orthogonally projected onto a surface. If the surface in Figure 12A is planar, rotations about both vertical and horizontal axes produce scale changes but preserve both collinearity and relative spacing in any

direction, as shown in Figure 12B. The position of any given point on this surface can be specified by second-order spatial derivatives in any direction. If the surface in Figure 12A is spherical, however, then 3-D rotation produces the image in Figure 12C—where both collinearities and interpoint distances have been disrupted. For this curved surface, the point's 3-D surface position entails fourth-order image structure, involving conjoint measures of second-order relations in multiple directions. Rigid 3-D motions of arbitrarily curved surfaces alter second-order but not fourth-order image structure.

Observers were required in Experiment 3 to adjust the position of a target point on a spherical as well as a planar surface. Additional comparisons also were provided by simpler three- and five-point patterns, testing acuity for positions defined by second-order relations of quasi-collinearity and differential compression, similar to those used by Lappin et al. (1995).

Method

Stimuli and tasks. The five stimulus patterns are illustrated schematically in Figure 13. A curved hexagonal pattern, which is the pattern of principal concern, consisted of a 19-point hexagonal array orthogonally projected onto an invisible spherical surface of radius 2.5 cm (2.5°). This hexagonal array assumed one of six possible orientations in the image, rotated in two dimensions between 0° and 50° about the viewing axis. The pattern was also slanted in depth relative to the image plane, first rotated 20° about the horizontal axis and then 20° about the vertical axis, so that the surface normal in the center of the surface pointed slightly up and to the left. This standing rotation of the curved hexagonal pattern meant that the correct 3-D location of the central target point produced an image position that was neither collinear with nor centered between other points (see Figures 12 and 13). The 3-D shape of the surface was stereoscopically defined in the stereo conditions (5.9 arcmin disparity between center and nearest points; 13.6 arcmin from center to the outer points) and by equivalent two-frame alternating rotations ($\pm 3^\circ$) about the vertical axis in the motion conditions. The hexagonal array used for this spherically curved pattern was created so that in the rotated position, the average image separation between the central target point and the immediately surrounding six points was 1°. As in the previous experiments, the target point adjusted by the observer was at the center of the hexagonal array.

The planar hexagonal pattern was similar to the curved pattern, but the 19-point array was projected onto a planar surface. As indicated above, the image pattern was randomly rotated 0° to 50° in two dimensions around the view axis. To approximate the image-to-image deformations of the curved pattern, we applied horizontal expansions and contractions (by $\pm 1.91\%$) dichoptically to the two stereoscopic half-images and in an alternating two-frame sequence in the motion conditions. Further, to match the first-order changes in image orientation produced by rotating the curved patterns in depth, dichoptic 2-D image rotations ($\pm 1.02^\circ$) were applied to the stereoscopic half-images and in alternating two-frame sequences in the motion conditions. In the stereoscopic conditions, the planar surface appeared to be slanted in depth comparably to the apparent orientation of the curved hexagonal pattern, but in the motion conditions, the alternating expansions, contractions, and rotations appeared only as 2-D scale changes rather than 3-D rotations.

Three simpler patterns were also examined: A quasi-collinear (QC) pattern consisted of three vertically arranged collinear points separated by 1.0°—involving collinearity in the perceived horizontal position, depth, and relative motion of the central target point. Similarly, three horizontally aligned points in the differential compression (DC) pattern provided information involving differential compression of the interpoint separations. The QC-DC pattern combined both the QC and DC properties in a five-point 2-D cross pattern, as if the QC and DC patterns were superim-

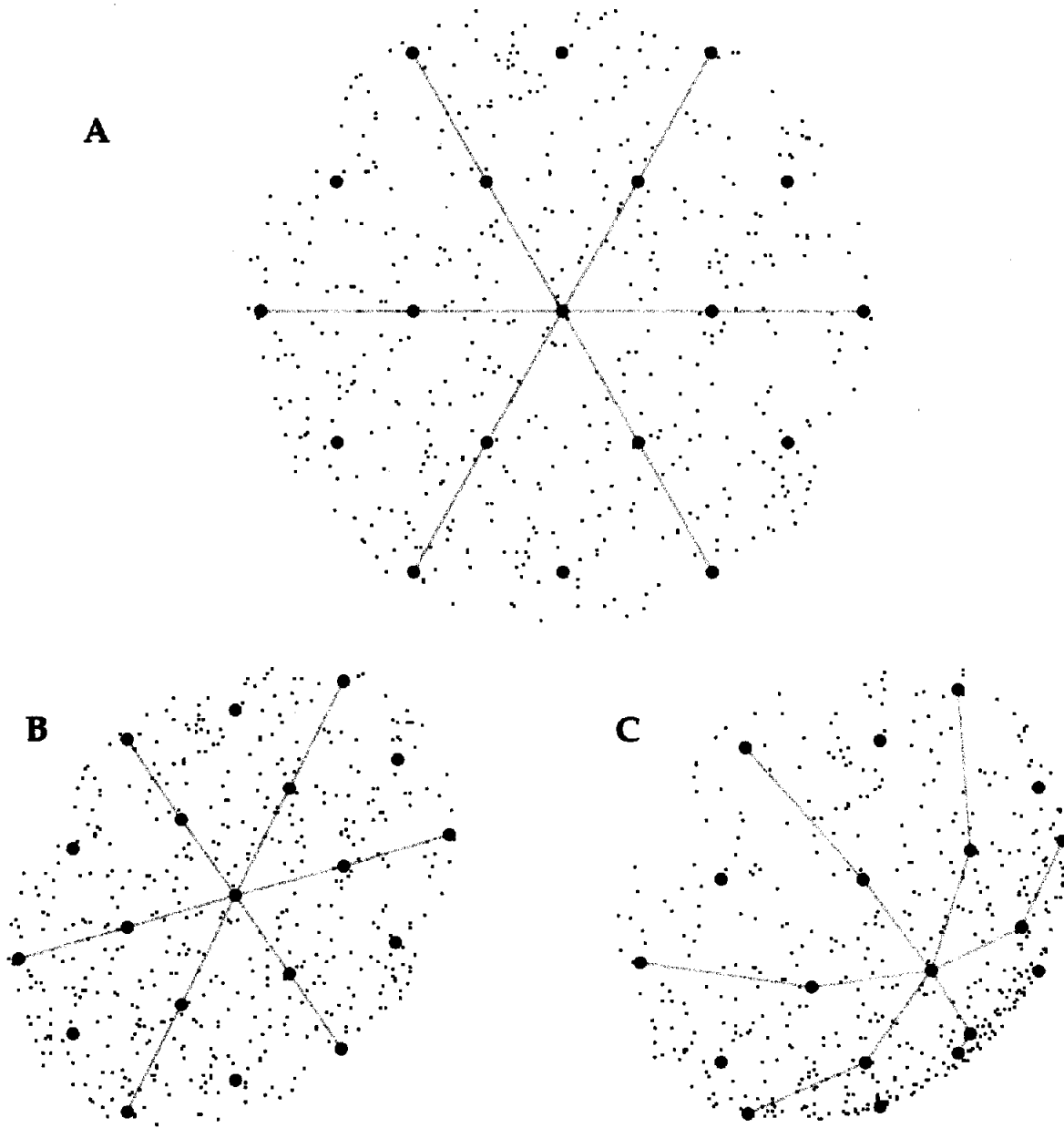


Figure 12. Illustration of second-order relationships on planar and curved surfaces. A: Hexagonal array orthogonally projected onto a planar or spherical surface and viewed along the axis of projection. The gray lines serve to emphasize the collinear relations and constant interpoint distances in the image. B: The planar array is tilted down and to the right. Note that all collinear relations are preserved and ratios of interpoint distances in any given direction remain constant in the image. C: The curved surface is now tilted down and to the right. Note that the collinear relationships are disrupted and interpoint distances have undergone differential compressions across the image. For this curved surface, maintenance of a point's 3-D surface position requires more complex image relations, involving both quasi-collinearity and differential compression. Rigid object motions entail changes in second-order spatial relations in the image. Invariance of the surface shape under arbitrary 3-D motions is represented in fourth-order image structure.

posed. The same interimage expansions, contractions, and 2-D rotations applied to the planar pattern were also applied to these three simpler patterns.

As in Experiments 1 and 2, lower order information about the target position was disrupted by random image transformations that perturbed

both the absolute positions and the lower order relations between the target and surrounding points. Zero-, first-, and second-order relations were disrupted by slanting the curved hexagonal pattern in depth as described above. Thus, when this 3-D shape was displayed stereoscopically or by rotation in depth, the position of the target point on the surface was defined

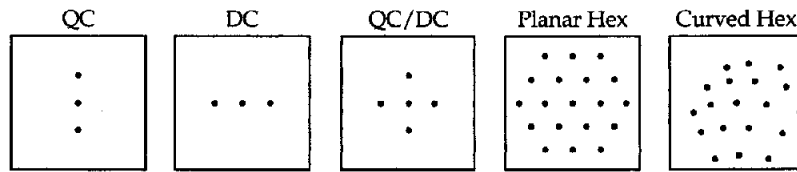


Figure 13. Schematic illustrations of the five stimulus patterns in Experiment 3. QC = quasi-collinear; DC = differential compression; Hex = hexagonal array.

only by the relative magnitudes of the changes in quasi-collinearity and differential compression and not by either of these second-order properties alone.

For both the planar hexagon and QC–DC patterns, either or both of the quasi-collinearity and differential compression relations could be used independently to detect the relative position, depth, and motion of the target point. For the QC pattern, this information was provided only by changes in quasi-collinearity. For the DC pattern, this spatial information was provided only by differential compression. The expansions and rotations imposed on all four of the planar patterns disrupted the first-order relations among points; neither position, depth, nor relative motion could be determined on the basis of orientations or separations between pairs of points.

In addition, information about absolute retinal positions was perturbed by random translations. All patterns underwent globally rigid random 2-D translations, and in the stereoscopic conditions these random translations were binocularly uncorrelated. The magnitude of this jitter was the same as in Experiment 2.

For the four planar patterns, the stereoscopic and motion detection tasks were the same as in Experiment 2, requiring stereo positioning of the target onto a planar surface slanted in depth or eliminating of nonrigid motion of the target relative to the surface. As before, the observers simultaneously performed the bisection task, centering the target horizontally between the surrounding points. For the curved hexagonal pattern, the task was conceptually the same but now involved the implicitly defined curved surface: Observers adjusted the stereo position of the target to lie on the spherical surface or tried to eliminate its motion relative to the surrounding surface. The bisection task for this spherical surface also required centering the target horizontally between the surrounding points on the surface in depth.

Procedure. The procedure was similar to that in Experiments 1 and 2. Stereo and motion sessions were interleaved in counterbalanced order until four sessions of each were completed. Each session consisted of four blocks of trials, with five trials per block, one block for each of four different stimulus configurations—the three- and five-point planar patterns and the 19-point curved hexagonal pattern.

Because the obtained acuities for the curved hexagonal pattern were surprisingly accurate, we later thought this might have resulted partly from potentially redundant information in these 19-point patterns. Accordingly, the planar hexagonal pattern was added as a more appropriate 2-D control pattern. The planar hexagonal pattern was tested in two separate experimental sessions, one for the stereo and one for the motion conditions, shortly after the other eight sessions.

Results and Discussion

The bisection and the stereoscopic and motion detection acuities for each pattern are shown in Figure 14. The most important comparison is between the planar and curved hexagonal patterns. The rms averages for these two patterns are shown in the top panel of Figure 14, and the lower panels show the more detailed results for each of the five patterns for each of the three observers.

As indicated by comparing the hyperacuities for the curved and planar hexagonal patterns, the visual surface position was robust

under disruptions of 1-D second-order relations. The average stereo acuity was 12.1 arcsec for the planar pattern and 15.4 arcsec for the curved pattern. The average motion detection acuity was 18.6 arcsec for the planar pattern and 20.4 arcsec for the curved pattern. Evidently, hyperacuity for relative position is based on fourth-order structure, involving a 2-D coordination of second-order relations.

The bisection acuities were significantly worse for the curved than for the planar patterns. Apparently, the bisection task depended on different geometric properties than the two hyperacuity tasks—probably involving comparisons of metric distances on either side of the target point, rather than visual stability of surface positions in the stereoscopic and motion detection tasks. The poorer bisection acuity for the curved than the planar hexagonal patterns indicates that observers were less sensitive to the relative distances over these curved surfaces than to the 2-D distances on the planar surfaces. A related fact is that the bisection acuity for differential compression in the DC patterns was noticeably worse than that for quasi-collinearity in the QC and QC–DC patterns. The curved hexagonal pattern, however, did not provide sufficient collinearity for centering the target point, because the 2-D projection of a geodesic on the curved surface is not a straight line, and a perceived target position relative to this geodesic would require perception of the metric curvedness of the surface. The present results are compatible with those of Perotti, Todd, Lappin, and Phillips (1998) in showing that observers do not obtain reliable information about the metric curvedness of surfaces, even though they are quite sensitive to the qualitative shape of the surface defined by the ratio of the two principal curvatures.

The results of Experiment 3 also clarify the results of Experiment 2 as well as those of Lappin et al. (1995), where acuity for detecting relative motion was not maintained under random scale changes in the horizontal axis. When the differential compression was combined with the second-order structure in a second dimension, as occurred in the curved pattern, then the perceived structure from motion benefited from this more complex fourth-order structure associated with the local surface shape.

The present findings are also consistent with the hypothesis that humans are limited to affine structure along the line of sight (Koenderink & van Doorn, 1991; Todd & Bressan, 1990). Affine transformations of the depth scale preserve the fourth-order differential structure associated with qualitative local surface shape.

General Discussion

Spatial Information for Vision Involves Correspondences Between Objects, Images, and Perceptions

This article is based on a generalized concept of information and offers a framework for identifying the spatial information for

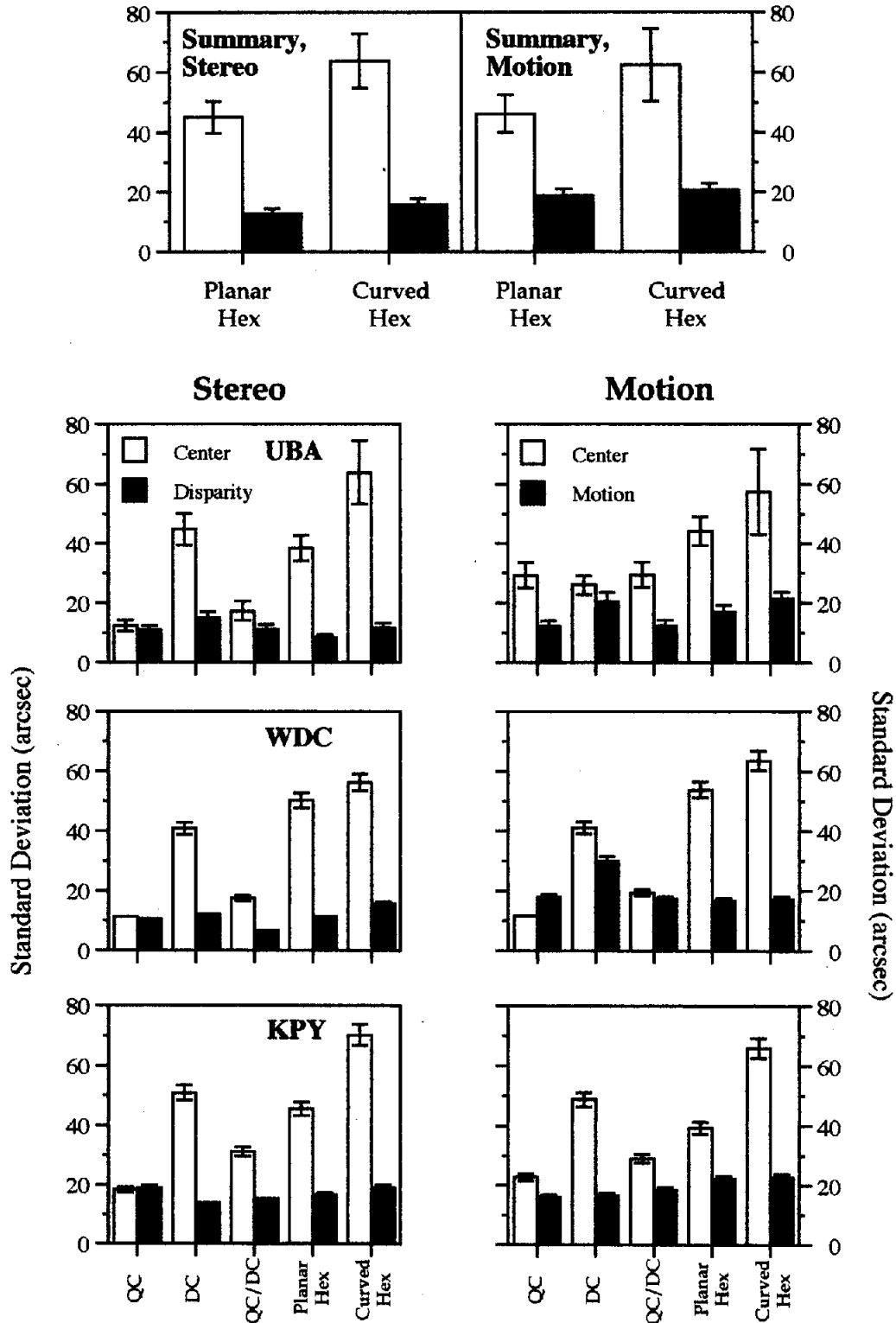


Figure 14. Results of Experiment 3, plotting positional variability versus pattern configuration under both stereo and motion conditions. The top two panels summarize average data (root-mean-square average for the three observers [UBA, WDC, and KPY]) for only the planar and curved hexagonal patterns (Hex), as these two conditions are of primary interest. The standard deviations of the stereo and motion detection adjustments were about 2–3 arcsec larger for the curved than for the planar hexagonal pattern, but the bisection adjustments were about 17–18 arcsec larger for the curved than for the planar hexagonal pattern. The bottom three rows of graphs show the data for each of the three observers, including data for the quasi-collinear (QC), differential compression (DC), and combined QC–DC patterns. Error bars designate ± 1 SE, estimated by a bootstrap method.

vision. In particular, visual information for perceiving objects consists of reliable spatial correspondences between environmental objects, their retinal images, and their perceptions. The general theoretical problem of spatial vision, therefore, is to identify spatial relations that are preserved in the mappings between these three domains.

The correspondence between environmental surfaces and their retinal images has been described mainly by Koenderink and van Doorn (1992a). The principal idea is Proposition 5, which identifies an isomorphism between local surface shape and retinal images. This isomorphism connects the fourth-order differential structures (interdependent second-order differentials in two dimensions) of environmental surfaces and of the displacement fields of their stereoscopic and moving images. This isomorphism does not exist, however, for lower order properties involving 1-D curvature, surface gradients or 3-D orientation, or distances between points. Thus, the correspondence of fourth-order differential structure of surfaces and their images is necessary for perceiving the shapes of environmental objects.

Whether or not vision is directly sensitive to this optical information is a logically separate question, although vision must be somehow represent this spatial structure. The present experimental results support the hypothesis (H4) that the spatial primitives of visual stimulation involve the differential image structure that identifies local surface shape.

Evidence that surface structure plays a fundamental role in visual perception has emerged from many lines of recent research—on stereopsis (Anderson & Nakayama, 1994; Lappin & Craft, 1997; Mallot, Arndt, & Bülhoff, 1995; Nakayama & Shimojo, 1992; Rogers & Cagenello, 1989; van Ee & Erkelens, 1996), 3-D structure from motion (Koenderink & van Doorn, 1992b; Norman & Lappin, 1992; Perotti et al., 1998; Rogers & Graham, 1983; Todd & Norman, 1991; Turner, Braunstein, & Andersen, 1995), shape perception (Koenderink, Kappers, Todd, Norman, & Phillips, 1996; Phillips, Todd, Koenderink, & Kappers, 1997), and brightness perception (Adelson, 1993; Kanizsa, 1979). The present results extend this evidence by showing that hyperacuties for detecting motion and binocular disparity also involve perception of surface shape.

Less well-known, perhaps, is the related result that discriminations of lower order properties such as surface slant, depth, and metric structure often are inaccurate and inconsistent (e.g., Norman & Todd, 1996, 1998; Norman, Todd, Perotti, & Tittle, 1996; Perotti et al., 1998; Proffitt, Bhalla, Gossweiler, & Midgett, 1995; Reichel, Todd, & Yilmaz, 1995). On the other hand, when local surface orientations are judged at many locations on a single smooth surface, then judgments by a given observer usually are highly consistent with one another (Koenderink et al., 1996; Todd, Norman, Koenderink, & Kappers, 1997). Inaccuracies and inconsistencies seem to arise mainly in variations between objects, observers, and viewing conditions, and they often involve the global depth scale for the surface as a whole. The qualitative shape of the surface, however, usually is perceived very accurately and reliably, even when discriminations of lower order properties are unreliable (Perotti et al., 1998). Thus, the present results converge with those of other recent studies in showing that vision detects accurate and reliable information about qualitative local surface shape but not about the lower order spatial properties of surfaces.

Perceived surface shape must not be derived from perceptions of lower order primitives.

Surface perception often is thought to emerge relatively late in the neural stream of visual processing as a product of cortical mechanisms. Cortical mechanisms doubtless play a critically important role, but they cannot be the origin of this spatial structure. Evidently, the spatial structure of surfaces is represented by the retina!

Spatial Primitives and Hyperacuity

As summarized in Proposition 9, hyperacuity represents an upper bound on the spatial uncertainty of retinal stimulation. Spatial uncertainty can only increase with subsequent neural operations. Hyperacuity, therefore, offers a powerful methodological tool.

The conclusion that spatial primitives of vision involve fourth-order differential structure associated with local surface shape is based on the finding that discriminations of these higher order spatial relations are robust under the group of rigid motions in 3-D space. These 3-D motions involved random variations in the component zero-, first-, and second-order relations. If visual information about the fourth order was derived from lower order primitives, then acuity thresholds for the higher order structure necessarily would be greater than those for lower order structure and would increase substantially with the noise in lower order relations.

The quantitative as well as qualitative similarities between the stereoscopic and motion hyperacuties provide converging evidence that the spatial inputs to both of these visual processes are similar. The one exception to this converging evidence is the discrepancy between the stereoscopic and motion detections of the 1-D second-order differential compression property, which seems to be poorly detected by the motion system. In other respects, hyperacuties for both binocular disparity and relative motion seem to be based on the fourth-order spatial structure that identifies local surface shape.

Visual Mechanisms

The present experiments were not designed to test specific hypotheses about the visual mechanisms underlying acuties for relative spatial position, but they raise questions about such mechanisms. The following discussion is necessarily speculative, partly because currently available physiological evidence is incomplete. Nevertheless, several intuitive ideas about the physiology of spatial vision warrant reexamination.

Revised concepts of visual representations of spatial relations. The idea that retinal mechanisms represent the higher order differential structure of stimulation seems to demand at least three significant departures from the usual conceptions of spatial vision mechanisms.

First, spatial positions and relationships must be tied to the intrinsic optical structure rather than to extrinsic locations on the retina or visual anatomy. The conventional idea is that the spatial structure of optical stimulation is represented through the apertures of photoreceptors and neurons whose spatial position are specified independently of the stimulation. Such an extrinsic framework is problematic for representing the spatial structure of optical pat-

terns, however, and it is challenged by the present results. When stimulus patterns move over the eye, physiological representations of their spatial structure must preserve information about the spatial arrangement of extrema, curvature inflections, and other features.

A basic idea in the present theoretical development is that stimulus structure is described by its differentials over retinal space, by terms of the form $\partial r/\partial s$, $\partial^2 r/\partial s^2$, $(\partial^2 r_i/\partial^2 r_j)/(\partial s_i^2/\partial s_j^2)$, and so forth, where r_i and s_j designate coordinate positions on the retina and on the object surface, respectively. Retinal coordinates are relevant in this formulation, but they do not define the spatial positions and relations, and they need satisfy only weak structural properties. For representing the differential structure of stimulation, the retinal coordinates need only homogeneity from one retinal neighborhood to another, so that the relative magnitudes of the second-order partial derivatives, $\partial^2 r/\partial s^2$, are not altered by small translations and rotations (of a few minutes of arc). No assumptions are needed about the scaling of these spatial derivatives, nor about metric units in the retinal coordinates, nor about orthogonality of axes.

Second, physiological representations must preserve spatiotemporal relations, with measures at one point correlated with those at other spatially separate points. The observed hyperacuties for relative positions of spatially separate features seem to require coherence (correlations of stochastic errors) among responses of separate cells. Both the structure and function of the early spatial mechanisms seem to involve multicell assemblies, spatially extended and relational in nature, not embodied in individual neurons or receptors.

Third, in contrast to the sometimes implicit assumption that retinal coordinates or other visual spatial mechanisms represent metric spatial relations, physiological representations of spatial relations probably are nonmetric. The results especially of Experiment 3 indicate that early visual information about spatial positions and relations is shaped by the surrounding stimulation so as to preserve information about local surface shape, involving 3-D relations rather than only 2-D retinal distances. That is, visual space seems to be derived as a representation of the intrinsic structure of the stimulation rather than given by an a priori reference frame.

Representing fourth-order image structure. With these ideas in mind, how can higher order differential structure be measured directly, without comparing more elementary measures of lower order structure? Consider, for example, how one might measure first-order differential relations associated with pairwise lengths and orientations without comparing zero-order measures of retinal positions. Such first-order measures are provided by the center-surround organization of neural receptive fields found in retinal ganglion cells and other visual areas of many species. The zero-order positions of these receptive fields are neither necessary for representing the first-order relation nor indicated by available physiological evidence. Center-surround mechanisms will not suffice for representing second-order relations, but they indicate how higher order spatial relations need not be derived by comparing retinal positions.

The same approach can also be extended to second- and fourth-order relations, by constructing local receptive-field operators from derivatives of a Gaussian kernel as discussed by Koenderink and van Doorn (1992b) and Koenderink (1990). This method

measures higher order derivatives by integrating the local image, and it does not introduce prohibitive levels of noise associated with independent measures of lower order primitives.

Such univariate receptive fields do not suffice for the present results, however, because the necessary spatial output is inherently multidimensional, involving multiple degrees of freedom. For a fixed direction, measures of second-order differentials require only one degree of freedom; independent (2-D) measures of the two first-order components are unnecessary and more vulnerable to noise. When such differentials must be measured in multiple spatial directions relative to a surface whose directions of principal curvature are not fixed beforehand, at least three directions of measurement are required. The fourth-order spatial primitives for representing local surface shape, as illustrated in Figures 3 and 4, require three degrees of freedom to measure relative curvatures in two principal directions and also specify these directions relative to the retina (see Koenderink & van Doorn, 1992a).

As discussed by Lappin and Craft (1997), second-order differential relations might be represented by phase relations (Fleet, Wagner, & Heeger, 1996; Freeman & Ohzawa, 1990; Smallman & MacLeod, 1994). Cortical mechanisms like those described by Freeman and Ohzawa and by Fleet et al. might well be involved in the phenomena documented in the present study, but these appear insufficient, for two reasons. First, hyperacuties for second- and fourth-order spatial relations appear to involve retinal phase relations, because the obtained hyperacuity for these relationships could not be recovered at the cortex if it were not available in the retinal signals. Second, local fourth-order structure entails three degrees of freedom, requiring multiple cells for its representation rather than univariate responses of individual neurons.

Hyperacuity from multiple images. The hyperacuties obtained in this study involve remarkably precise information about extended spatial relations. Evidently, such hyperacuties are associated with broader aspects of visual perception than usually recognized.

Understanding how visual mechanisms achieve these hyperacuties for binocular disparity and relative motion is a challenging problem, however. One salient question is how these multi-image hyperacuties are so much better than the bisection acuties for static nonstereoscopic images. The static monocular images must serve in some sense as the input to the binocular and motion systems and would seem to limit acuties for binocular disparity and relative motion. This problem has also been considered by previous authors (Lappin & Craft, 1997; McKee et al., 1990; Westheimer, 1979a; Westheimer & McKee, 1979), but a satisfying explanation is not yet available. Two observations are pertinent, however.

First, the finding in Experiment 1 that Weber fractions for the bisection thresholds were about five times greater than those for the stereoscopic and motion hyperacuties indicates that fine-grained spatial structures of stationary monocular images are visually correlated between the two eyes and over time. The same conclusion about stereopsis was reached by Legge and Gu (1989) on the basis of stereoscopic and monocular contrast sensitivities. Their estimate of .98 for the binocular correlation was essentially the same as that obtained by Lappin and Craft (1997) for stereoacuity and was also very similar to the correlations implied by the results of Experiment 1 for both stereopsis and relative motion. The binocular correlations for the stereo detection and bisection judgments are illustrated in Figure 15 for observer WDC in Ex-

Observer WDC

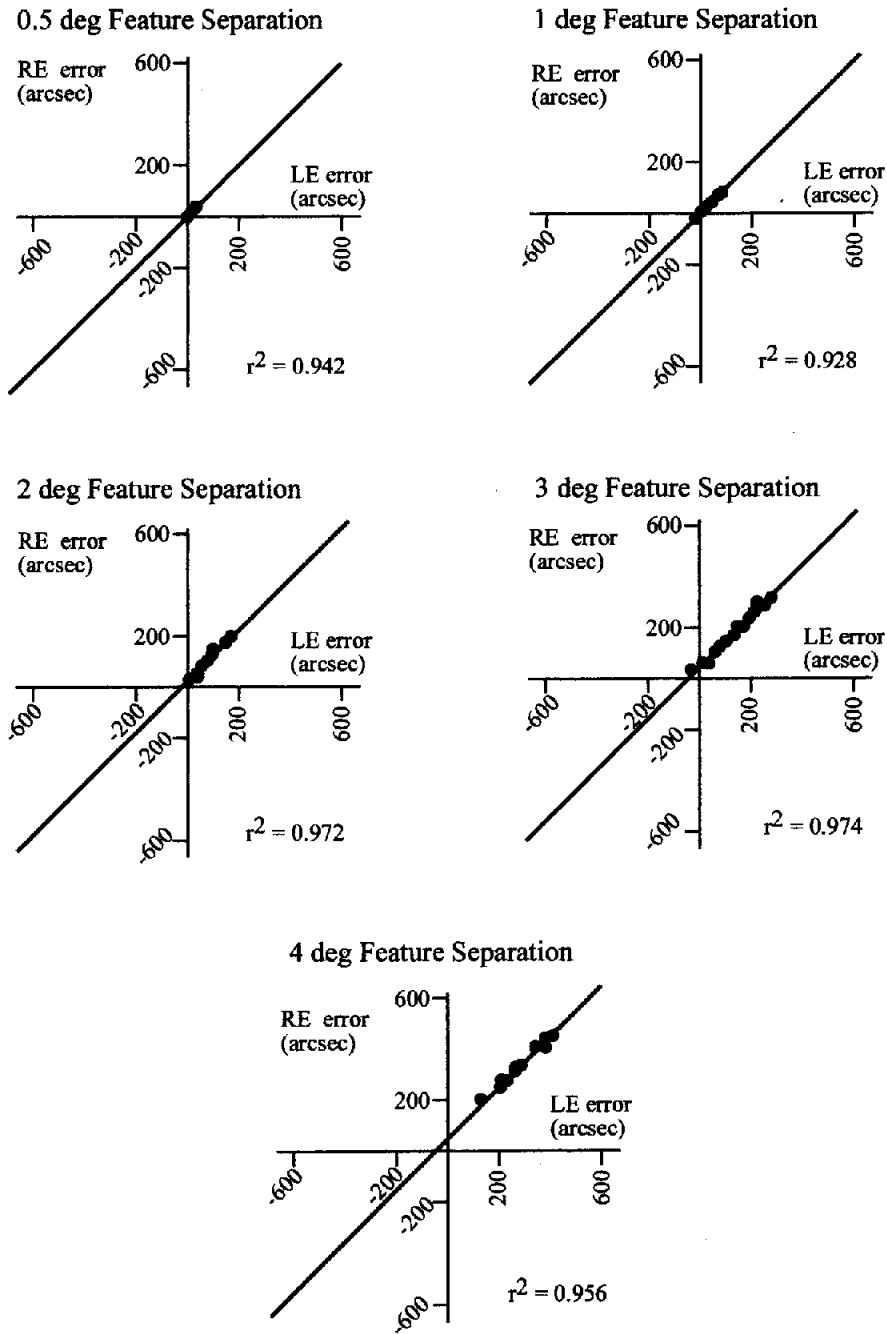


Figure 15. Scatterplots showing the distributions of the adjusted target positions in the left and right images in the stereo task of Experiment 1 for one observer. Results for the other two observers were qualitatively and quantitatively similar. deg = degree; RE = right eye; LE = left eye.

periment 1. (Data for the other two observers were quantitatively very similar.) The data in Figure 15 help in visualizing the relationship between the two monocular spatial positions, though the correlations are already implied by the superiority of the stereoscopic hyperacuties relative to the concurrent bisection acuties.

These scatterplots do not explain how such interocular correlations might be achieved.

Very similar estimates for the temporal coherence of moving images also may be obtained for the motion hyperacuties. Scatterplots like those in Figure 15 are not available for the motion

data, but one can estimate the correlation from the obtained standard deviations of the bisection and relative motion adjustment responses, S_B and S_M , respectively:

$$r = (S_{B1}^2 + S_{B2}^2 - S_{B1-B2}^2)/(2S_{B1} S_{B2}) = (2S_B^2 - S_M^2)/(2S_B^2), \quad (5)$$

where S_B is estimated by the standard deviations of bisection judgments in each of two successive frames, say $S_B = S_{B1} = S_{B2}$, and S_M is estimated by the variability of the visual difference in positions between successive frames, say $S_M = S_{B1-B2}$. All 15 of the temporal correlations estimated in this way from the motion hyperacuties of the three observers for the five feature separation conditions in Experiment 1 were above .97. These similar estimated binocular and temporal coherence values are striking and intriguing, and they seem to point to physiological characteristics that merit further study.

The hypothesis that retinal spatial signals are coherent over time and between the two eyes suggests a possible mechanism for the remarkable precision of these hyperacuties. Lappin and Craft (1997) discussed the possibility that stereoacuity might involve phase interference between two coherent images, analogous to the spatial resolution obtained by superimposing coherent images in holography, radio astronomy, and side-looking radar (Gabor, 1966; Readhead, 1982). As Lappin and Craft observed, such hyperacuties could occur if physiological responses had a wavelength of about twice the separation between foveal cones and if binocular and moving images were highly correlated.

Without superimposing such images by stereopsis and relative movement, the spatial information latent in these images would not be observable; spatial positions would be signaled only by the amplitude envelope of the retinal images. The optical point-spread function of the eye has a diameter of about 2 arcmin, so that two-point resolution thresholds are limited to about 1 arcmin (see Geisler, 1989). If binocular and motion mechanisms simply added or subtracted such images, then the same limits would apply to these physiological combinations. The empirical result that hyperacuties for binocular disparity and differential motion achieve much better resolution implies that retinal spatial signals have a much finer structure. Previously, hyperacuties for disparity and motion have been attributed to local retinal signs (e.g., McKee et al., 1990). The present results, however, show that these spatial signals derive from intrinsic relations among the optical features rather than extrinsic local signs. The hypothesis that hyperacuties for disparity and motion derive from coherent retinal responses distributed among separate features is speculative, but alternative hypotheses are not now known.

If vision provides precise spatial information about the relative positions of separate features, then why is this less effective for bisection judgments in the present and similar previous experiments (e.g., K. K. De Valois, Lakshminarayanan, Nygaard, Schlüssel, & Sladky, 1990)? The two tasks evidently rely on different spatial information. Judgments of relative position in single images seem to require information about relative distances between features, but hyperacuties for disparity and motion probably involve only detection of changes in relative position; information about distances between features is not required for these hyperacuties. Perhaps the visual system does not preserve information about retinal distances between features, partly because these are so poorly correlated with environmental spatial relations.

Anatomical and physiological substrate. How could the retina provide information about spatial relations among optical features separated by several degrees, with a precision of about 0.25% of the separation? The present results indicate that the spatial signals from separate retinal areas must be correlated, with a correlation length of at least 4°. How might this be possible? A plausible hypothesis is that spatial coherence is enabled by the extensive network of horizontal and amacrine cells (see Rodieck, 1998). Each photoreceptor contacts many horizontal cells, each horizontal cell contacts many receptors, and horizontal cells are electrically connected by gap junctions within rods and cones. Both horizontal cells and amacrine cells have dense dendritic arbors and axons of several millimeters. Horizontal cells are known to provide lateral inhibition among photoreceptors, and amacrine cells provide both inhibition and excitation. These lateral interactions might also serve to maintain coherent spatiotemporal phase relations among signals from separate retinal locations.

Summary

1. Vision depends on one-to-one correspondences of spatial structure in environment, retina, brain, and perception.
2. Fourth-order differential structure of environmental surfaces is isomorphic with that of interimage displacement fields, constituting information about local surface shape. Lower order spatial structure does not provide this information.
3. Hyperacuties for binocular disparity and relative motion were found to be based on the intrinsic optical image structure rather than on extrinsic retinal coordinates. These hyperacuties were robust under random 3-D motions that perturbed lower order spatial structure, so the spatial primitives must involve the higher order structure.
4. Spatial information about local surface shape must be represented by retinal signals. The retinal signals from separate features seem to be coherent over several degrees of arc.
5. The underlying visual mechanisms require further investigation.
6. The theoretical framework and experimental findings of this study help to explain the visual perception of environmental surface shape.

References

- Adelson, E. H. (1993, December 24). Perceptual organization and the judgment of brightness. *Science*, 262, 2042-2044.
- Adelson, E. H., & Bergen, J. R. (1991). The plenoptic function and the elements of early vision. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 3-20). Cambridge, MA: MIT Press.
- Anderson, B. L., & Nakayama, K. (1994). Toward a general theory of stereopsis: Binocular matching, occluding contours, and fusion. *Psychological Review*, 101, 414-445.
- Berry, R. N. (1948). Quantitative relations among vernier, real depth, and stereoscopic depth acuities. *Journal of Experimental Psychology*, 38, 708-721.
- Brindley, G. S. (1970). *Physiology of the retina and visual pathway*. Baltimore: Williams & Wilkins.
- Collewijn, H., & Erkelens, C. J. (1990). Binocular eye movements and the perception of depth. In E. Kowler (Ed.), *Eye movements and their role in visual and cognitive processes* (pp. 213-261). Amsterdam: Elsevier Science.

- den Dekker, A. J., & van den Bos, A. (1997). Resolution: A survey. *Journal of the Optical Society of America A*, *14*, 547–557.
- De Valois, K. K., Lakshminarayanan, V., Nygaard, R., Schlussel, S., & Sladky, J. (1990). Discrimination of relative spatial position. *Vision Research*, *30*, 1649–1660.
- De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision*. New York: Oxford University Press.
- 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.
- Doner, J., Lappin, J. S., & Perfetto, G. (1984). Detection of three-dimensional structure in moving optical patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 1–11.
- Efron, B., & Tibshirani, R. (1991, July 26). Statistical data analysis in the computer age. *Science*, *253*, 390–395.
- Falzett, M., & Lappin, J. S. (1983). Detection of visual forms in space and time. *Vision Research*, *23*, 181–189.
- Fleet, D. J., Wagner, H., & Heeger, D. J. (1996). Neural encoding of binocular disparity: Energy models, positions shifts and phase shifts. *Vision Research*, *36*, 1839–1857.
- Foley, J. M. (1980). Binocular distance perception. *Psychological Review*, *87*, 411–434.
- Freeman, R. D., & Ohzawa, I. (1990). On the neurophysiological organization of binocular vision. *Vision Research*, *30*, 1661–1676.
- Gabor, D. (1966). Holography, or the “whole picture.” *New Scientist*, *13*, 74–78.
- Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*, 267–314.
- Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Hernstein, R. J., & Boring, E. G. (Eds.). (1965). *A source book in the history of psychology*. Cambridge, MA: Harvard University Press.
- Horn, B. K. P., & Brooks, M. J. (Eds.). (1989). *Shape from shading*. Cambridge, MA: MIT Press.
- Johansson, G. (1994). Perception of motion and changing form. In G. Jansson, S. S. Bergström, & W. Epstein (Eds.), *Perceiving objects and events* (pp. 134–168). Hillsdale, NJ: Erlbaum. (Original work published 1964)
- Johnston, A., & Passmore, P. J. (1994a). Shape from shading: I. Surface curvature and orientation. *Perception*, *23*, 169–189.
- Johnston, A., & Passmore, P. J. (1994b). Shape from shading: II. Geodesic bisection and alignment. *Perception*, *23*, 191–200.
- Julesz, B. (1971). *Foundations of cyclopean vision*. Chicago: University of Chicago Press.
- Kanizsa, G. (1979). *Organization in vision*. New York: Praeger.
- Klein, S. A., & Levi, D. M. (1985). Hyperacuity thresholds of 1 second: Theoretical predictions and empirical validation. *Journal of the Optical Society of America A*, *2*, 1170–1190.
- Koenderink, J. J. (1986). Optic flow. *Vision Research*, *26*, 161–179.
- Koenderink, J. J. (1987). An internal representation of solid shape based on the topological properties of the apparent contour. In W. Richards & S. Ullman (Eds.), *Image understanding 1985–1986* (pp. 257–285). Norwood, NJ: Ablex.
- Koenderink, J. J. (1990). *Solid shape*. Cambridge, MA: MIT Press.
- Koenderink, J. J., Kappers, A. M. L., Todd, J. T., Norman, J. F., & Phillips, F. (1996). Surface and attitude probing in stereoscopically presented dynamic scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 869–878.
- Koenderink, J. J., & van Doorn, A. J. (1975). Invariant properties of the motion parallax field due to the movement of rigid bodies relative to the observer. *Optica Acta*, *22*, 773–791.
- Koenderink, J. J., & van Doorn, A. J. (1976a). Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics*, *21*, 29–35.
- Koenderink, J. J., & van Doorn, A. J. (1976b). Local structure of movement parallax of the plane. *Journal of the Optical Society of America*, *66*, 717–723.
- 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). The shape of smooth objects and the way contours end. *Perception*, *11*, 129–137.
- Koenderink, J. J., & van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A*, *8*, 377–385.
- Koenderink, J. J., & van Doorn, A. J. (1992a). Second-order optic flow. *Journal of the Optical Society of America A*, *9*, 530–538.
- Koenderink, J. J., & van Doorn, A. J. (1992b). Generic neighborhood operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*, 597–605.
- Koenderink, J. J., & van Doorn, A. J. (1996). Illuminance texture due to surface mesostructure. *Journal of the Optical Society of America A*, *13*, 452–463.
- Koenderink, J. J., & van Doorn, A. J. (1997). The generic bilinear calibration-estimation problem. *International Journal of Computer Vision*, *23*, 217–234.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). New York: Academic Press.
- Lappin, J. S., & Ahlström, U. B. (1996, November). Hypersensitivity to deformations of surface shape in moving images. *Abstracts of the Psychonomic Society 37th Annual Meeting*, p. 53.
- Lappin, J. S., Ahlström, U. B., Craft, W. D., & Tschantz, S. T. (1995). Spatial primitives for seeing three-dimensional shape from motion. In T. V. Pappathomas, C. Chubb, A. Gorea, & E. Kowler (Eds.), *Early vision and beyond* (pp. 145–153). Cambridge, MA: MIT Press.
- Lappin, J. S., & Bell, H. H. (1976). The detection of coherence in moving random-dot patterns. *Vision Research*, *16*, 161–168.
- Lappin, J. S., & Craft, W. D. (1997). Definition and detection of binocular disparity. *Vision Research*, *37*, 2953–2974.
- Legge, G. E., & Gu, Y. (1989). Stereopsis and contrast. *Vision Research*, *29*, 989–1004.
- Mallot, H. A., Arndt, P. A., & Bülthoff, H. H. (1995). *The mechanism of intensity-based stereo* (Tech. Rep. No. 18). Tübingen, Germany: Max-Planck-Institut für biologische Kybernetik.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- McKee, S. P., Welch, L., Taylor, D. G., & Bowne, S. F. (1990). Finding the common bond: Stereoacuity and the other hyperacuties. *Vision Research*, *30*, 879–891.
- Mingolla, E., & Todd, J. T. (1986). Perception of solid shape from shading. *Biological Cybernetics*, *53*, 137–151.
- Morita, T., & Kanade, T. (1997). A sequential factorization method for recovering shape and motion from image streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*, 858–867.
- Nakayama, K., & Shimojo, S. (1992, September 4). Experiencing and perceiving surfaces. *Science*, *257*, 1357–1363.
- Norman, J. F., & Lappin, J. S. (1992). The detection of surface curvatures defined by optical motion. *Perception & Psychophysics*, *51*, 386–396.
- Norman, J. F., & Todd, J. T. (1996). The discriminability of local surface structure. *Perception*, *25*, 381–398.
- Norman, J. F., & Todd, J. T. (1998). The discrimination of interval and ordinal depth relations on smooth surfaces and in empty space. *Perception*, *27*, 257–272.
- 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.
- Nowakowski, J., & Elbaum, M. (1983). Fundamental limits in estimating

- light pattern position. *Journal of the Optical Society of America*, 73, 1744-1758.
- 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.
- Phillips, F., Todd, J. T., Koenderink, J. J., & Kappers, A. M. L. (1997). Perceptual localization of surface position. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 1481-1492.
- Proffitt, D. R., Bhalla, M., Gossweiler, R., & Midgett, J. (1995). Perceiving geographical slant. *Psychonomic Bulletin & Review*, 2, 409-428.
- Readhead, A. C. S. (1982, June). Radio-astronomy by very-long-baseline interferometry. *Scientific American*, 246, 52-61.
- Regan, D., Erkelens, C. J., & Collewijn, H. (1986). Necessary conditions for the perception of motion in depth. *Investigative Ophthalmology & Visual Science*, 27, 584-597.
- Reichel, F. D., Todd, J. T., & Yilmaz, E. (1995). Visual discrimination of local surface depth and orientation. *Perception & Psychophysics*, 57, 1233-1240.
- Roberts, F. S. (1979). *Measurement theory: Vol. 7. Encyclopedia of Mathematics and Its Applications* (G.-C. Rota, Series Ed.). Reading, MA: Addison-Wesley.
- Rodieke, R. W. (1998). *The first steps in seeing*. Sunderland, MA: Sinauer.
- Rogers, B. J., & Cagenello, R. (1989). Disparity curvature and the perception of three-dimensional surfaces. *Nature*, 339, 135-137.
- Rogers, B. J., & Graham, M. E. (1983, September 30). Anisotropies in the perception of three-dimensional surfaces. *Science*, 221, 1409-1411.
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- Smallman, H. S., & MacLeod, D. I. A. (1994). Size-disparity correlation in stereopsis at contrast threshold. *Journal of the Optical Society of America A*, 11, 2169-2183.
- Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics*, 48, 419-430.
- Todd, J. T., & Norman, J. F. (1991). The perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception & Psychophysics*, 50, 509-523.
- Todd, J. T., Norman, J. F., Koenderink, J. J., & Kappers, A. M. L. (1997). Effects of texture, illumination and surface reflectance on stereoscopic shape perception. *Perception*, 26, 807-822.
- Tomasi, C., & Kanade, T. (1992). Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9, 137-154.
- Turner, J., Braunstein, M. L., & Andersen, G. J. (1995). Detection of surfaces in structure from motion. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 809-821.
- Tyler, C. W., & Julesz, B. (1978). Binocular cross-correlation in time and space. *Vision Research*, 18, 101-105.
- van Doorn, A. J., Koenderink, J. J., & van de Grind, W. A. (1984). Limits in spatio-temporal correlation and the perception of visual movement. In A. J. van Doorn, W. A. van de Grind, & J. J. Koenderink (Eds.), *Limits in perception* (pp. 203-234). Utrecht, the Netherlands: VNU Science Press.
- van Ee, R., & Erkelens, C. J. (1996). Stability of binocular depth perception with moving head and eyes. *Vision Research*, 36, 3827-3842.
- Wallach, H., & O'Connell, D. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, 45, 205-217.
- Westheimer, G. (1975). Visual acuity and hyperacuity. *Investigative Ophthalmology*, 14, 570-572.
- Westheimer, G. (1977). Spatial frequency and light-spread descriptions of visual acuity and hyperacuity. *Journal of the Optical Society of America*, 67, 207-212.
- Westheimer, G. (1979a). Cooperative neural processes involved in stereoscopic acuity. *Experimental Brain Research*, 36, 585-597.
- Westheimer, G. (1979b). The spatial sense of the eye. Proctor Lecture. *Investigative Ophthalmology & Visual Science*, 18, 893-912.
- Westheimer, G., & McKee, S. P. (1978). Stereoscopic acuity for moving retinal images. *Journal of the Optical Society of America*, 68, 450-455.
- Westheimer, G., & McKee, S. P. (1979). What prior uniocular processing is necessary for stereopsis? *Investigative Ophthalmology & Visual Science*, 18, 614-621.
- Wiener, N. (1948). *Cybernetics, or control and communication in the animal and the machine*. Cambridge, MA: MIT Press.
- Wiener, N. (1954). *The human use of human beings: Cybernetics and society*. Garden City, NY: Doubleday.
- Wilson, H. R. (1980). Spatiotemporal characterization of a transient mechanism in the human visual system. *Vision Research*, 20, 443-452.

Received June 8, 1998

Revision received February 15, 1999

Accepted February 18, 1999 ■