

THE PERCEPTION OF GEOMETRICAL STRUCTURE FROM CONGRUENCE¹

Joseph S. Lappin
Vanderbilt University
Nashville, Tennessee

and

Thomas D. Wason
Allotech Inc.
Raleigh, North Carolina

1. INTRODUCTION

The principal function of vision is to measure the environment. As demonstrated by the coordination of motor actions with the positions and trajectories of moving objects in cluttered environments and by the rapid recognition of solid objects in varying contexts from changing perspectives, vision provides real-time information about the geometrical structure and location of environmental objects and events.

Information about the geometrical structure of scenes, objects, and motions may be visually acquired not only by the exploration of natural environments, but also from artificial, human-designed displays. Photographs, drawings, movies, computer graphics, and other such artificial 2-D displays are widely and effectively used tools for communicating information about spatial structures. Understanding the basis for the effectiveness of such tools poses a special theoretical challenge, because the trigonometric mapping from the 3-D structures and motions portrayed in these displays to the optical patterns on the observer's retinae differs from the perspective projections that normally hold for vision in natural environments. Cutting (1987) has recently discussed the theoretical difficulties posed by this discrepancy between the projective geometry of movies versus that of natural vision, and he has also provided experimental demonstrations of the abilities of humans to perceive 3-D structure in movies viewed "from the front row side aisle."

The purpose of this paper is to examine the geometric information provided by 2-D spatial displays. We propose that the geometry of this information is best understood not within the traditional framework of perspective trigonometry, but in terms of the structure of qualitative relations defined by *congruences* among *intrinsic* geometric relations in images of *surfaces*. The mathematical details of this theory of the geometry of vision are presented elsewhere (Lappin, in press); the present paper outlines the basic concepts of this geometrical theory.

¹Work on this paper and on related experimental and theoretical research was supported in part by a Small Business Innovative Research Grant from NASA to T. D. Wason, by NIH Grant EY-05926 to J. S. Lappin, and by the University Research in Residence Program of the Air Force Office of Scientific Research which enabled several extended visits by Lappin to Wright-Patterson Air Force Base. The mathematical ideas outlined in this paper have benefitted significantly from discussions with Jan Koenderink and Andrea van Doorn, State University of Utrecht, The Netherlands, and especially with John G. Ratcliffe, Dept. of Mathematics at Vanderbilt.

Traditionally, the structure of space—both the 3-D space of the environment and the 2-D space of the image—has been regarded as defined *a priori*, independently of the objects and motions contained within it. Indeed, the geometric structure of objects and motions is typically described by reference to extrinsic standards that define parallel and perpendicular directions and quantify relative magnitudes of distance extrinsic to the objects themselves.

When described in terms of this extrinsic framework, however, the geometry of vision is quite complicated: Metric² relations in the 2-D image plane cannot be isomorphic with metric relations in the 3-D environment; the perspective projection from 3-D spatial structures in the environment onto the 2-D image plane does not have a well-defined inverse. Therefore, the recovery of information about the geometric structures and locations of the environmental objects has often been thought to require supplementary information about the perspective position of the observer or about the structure and location of the objects. The 2-D optical images alone have seemed insufficient.

But the assumption that vision begins with an abstract structure of space as a prior standard for describing environmental objects begs the question. The basic problem of vision is to find a measurement structure for representing the spatial characteristics of observed scenes, objects, and events. Such a measurement structure is generally not given beforehand, but must be discovered in the organization of the empirical observations themselves.

2. INTRINSIC GEOMETRY OF SURFACES AND IMAGES

When described in terms of the intrinsic geometry of *surfaces*, the geometry of vision becomes much simpler. In the first place, the mapping of a visible region of an environmental surface onto its optical image is a mapping from one 2-D manifold onto another. The derivatives and singularities of the surface—its slopes, peaks and valleys, inflections, saddlepoints, and occluding edges—are isomorphic with the derivatives and singularities of the image. This is true for images described by gradients of texture, motion parallax, or stereoscopic disparity (Koenderink and van Doorn, 1975, 1976a,b,c, 1977). Although the isomorphism does not hold for images described by luminance gradients, partly because of the additional influence of the direction of illumination, it is still true that the intrinsic surface structure (in particular, the parabolic lines, which are inflections of curvature that separate regions of convexity and concavity) is systematically related to the differential structure of the image (Koenderink and van Doorn, 1980). Because the differential structures of the two manifolds are essentially isomorphic with one another, the ordinal topography of the visible region of an environmental surface is fully described and recoverable by its optical image.

Furthermore, the specific mapping between curves and forms on the environmental surface and their corresponding images on an observer's retina may be locally described simply by a *linear*

²The term metric is used in a conventional mathematical sense, referring in this context to measures of distance over a potentially curved surface. A relation $m(a,b)$ between two elements a and b is said to be a metric relation if it satisfies the following axioms for all elements a , b , and c :

positivity: $m(a,b) \geq 0$

symmetry: $m(a,b) = m(b,a)$

reflexivity: $m(a,a) = 0$

triangle inequality: $m(a,c) \leq m(a,b) + m(b,c)$.

coordinate transformation between the derivatives on the two manifolds. This linear approximation holds for "infinitely small" surface patches that may be locally approximated by a tangent plane at that location. This linear mapping of the surface onto its image also has a well-defined inverse. Accordingly, the local structure of the surface may be obtained from the local structure of its image by a linear coordinate transformation.

These simple relationships between the surface and its image involve the *derivatives* on the two manifolds. The linear transformation that best describes the relationship between these two manifolds at any given point is given by the partial derivatives of the two coordinate systems. Thus, if O^2 represents the 2-D manifold of the object surface, and if R^2 represents the 2-D manifold corresponding to the observer's retina, then the linear differential map $v: O^2 \rightarrow R^2$ is specified by the following Jacobian matrix of partial derivatives:

$$V = \begin{bmatrix} \partial r^1 / \partial o^1 & \partial r^1 / \partial o^2 \\ \partial r^2 / \partial o^1 & \partial r^2 / \partial o^2 \end{bmatrix}$$

Suppose that $[dO] = [do^1, do^2]$ is a 2×1 column vector that specifies an infinitesimal displacement on the surface in terms of two intrinsic coordinates on the object surface, and suppose that $[dR] = [dr^1, dr^2]$ is a corresponding description of the image of this vector in terms of the intrinsic coordinates of the retina. Then the transformation between these two coordinate systems produced by the optical projection from the object to its image on the retina is given by the linear equation

$$[dR] = V[dO]$$

and the inverse map is given by

$$[dO] = V^{-1}[dR]$$

where V is the Jacobian matrix given above. (The form of this equation is independent of the specific coordinate systems used to specify positions on the two manifolds. The coordinates need not intersect at right angles nor even be straight lines; they need only be differentiable and to provide a unique specification of each position on the manifold. The generality of this representation seems especially relevant to vision, where no specific coordinate system can be assumed beforehand for any particular environmental surface.³) The important point is that *the local structure of the retinal image of a given surface is described by this Jacobian matrix of partial derivatives, V .* The entries in this matrix vary as a function of position on the surface, with variations in the values of these entries reflecting variations in the orientation and curvature of the surface.

The same approach can also be used to describe the relationships with a third 2-D manifold associated, for example, with an intervening display image such as a movie or photograph. Suppose that I^2 represents the manifold of such an intervening image, that $a: O^2 \rightarrow I^2$ represents the

³For concreteness, we may assume that the coordinates reflect the spatial arrangement of the gradients and singularities of the surface—e.g., tending to run parallel and perpendicular to the gradients of curvature of the surface and to the boundary contours, corners, and parabolic lines (which separate structurally distinct regions). We need not assume that these coordinates have specific numerical values, only that they are differentiable and uniquely label every location on the surface.

differential map between these two manifolds, and that $b: I^2 \rightarrow R^2$ is the visual map from the display image onto the retinal manifold. Then, using the chain rule, the two successive maps can be combined by a composition of the two functions, $v = (b \cdot a): O^2 \rightarrow R^2$. Similarly, the coordinate transformation corresponding to this chain would be given by a linear equation of the following form:

$$[dR] = BA[d0],$$

where the matrix product $BA = V$ again provides a linear coordinate transformation functionally equivalent to the previous construction.

Representation of the *metric* structure of the surface requires an embedding of the 2-D manifold of the surface or its image into the 3-D manifold of Euclidean 3-space, E^3 . Suppose that $[dX] = [dx^1, dx^2, dx^3]$ is a 3×1 column vector giving the three orthogonal cartesian coordinates of an infinitesimal displacement on the object surface. Then the perspective coordinate embedding of the image of the surface into E^3 , $p: R^2 \rightarrow E^3$, is given by a linear coordinate transformation of the following form:

$$[dX] = PV[d0]$$

where P is a 3×2 matrix of partial derivatives, $P = [\partial x^k / \partial r^i]$, with $k = 1, 2, 3$ and $i = 1, 2$. Measures of metric relations require a quadratic expression similar to the Pythagorean formula for distance in E^3 . The metric tensor that provides the measure of distance on the surface is obtained by substituting from the above equation for the vector $[dX]$ in the Pythagorean formula:

$$\begin{aligned} ds^2 &= [dX]^T [dX] \\ &= [PV[d0]]^T PV[d0] \\ &= [d0]^T V^T P^T PV[d0] \\ &= [d0]^T V^T P^* V[d0], \end{aligned}$$

where $P^* = P^T P$ is a symmetric 2×2 matrix with quadratic entries of the form

$$P^* = [\sum_k (\partial x^k / \partial r^i) (\partial x^k / \partial r^j)].$$

Thus, the entries in this matrix provide a measure of squared distance on the object surface at a particular position on the retina corresponding to the image of the surface. The length of any arbitrary curve on the surface is obtained by integrating the quantities ds defined in the preceding equation at each position along the curve.

The three independent parameters of the matrix P^* are not given directly by a single stationary image of an isolated local surface patch. In certain special cases these perspective parameters and therefore the metric structure of the local surface patch are determined, up to a scalar, simply by the motion of the local patch. More generally, however, these perspective parameters must be derived from more global constraints on the image structure associated with the observer's position

and motion within the 3-D environment. In general, the perspective embedding of the image into E^3 is revealed by actual or implied motions of objects within the space.

3. METRIC STRUCTURE FROM CONGRUENCE

Although geometric relations are often described in terms of extrinsic coordinate systems in which directions and distances are defined *a priori*, it is important in many applications to derive the structure of space from more fundamental qualitative relationships among the objects and events contained within it. This was the case, for example, in the development of relativistic physics, where the symmetries of observations associated with the velocity of light and with gravitation were used to construct spaces in which the lawful relations among observed variables could be expressed in simpler and more general form (Einstein *et al.* 1923, 1952; Misner, Thorne, and Wheeler, 1973). The same strategy has also been employed in formulating the theoretical foundations of measurement (Krantz *et al.*, 1971; Luce, 1978; Luce and Narens, 1983). That is, symmetries of qualitative relations under various physical operations and under varying conditions of observation may often be used as a foundation for quantitative equations that describe empirical laws of nature.

Analogously, the geometry of vision may also rest upon the symmetries of intrinsic qualitative relations in the spatio-temporal optical images rather than on the prior metric structure of an extrinsic coordinate system. Because metric relations in the 3-D environment are not isomorphic with those in the 2-D image, and because the optical projections of environmental objects onto the retinae change with the perspective positions of the displays and observers, the extrinsic framework of space is neither constant nor readily accessible to vision. Instead, we hypothesize, the metric structure of environmental objects and spaces may be induced from the isometries of moving objects.

This conception of the geometry of vision is a continuation of ideas developed by Gibson (1950, 1957) about the importance of the concepts of invariance and transformations for perception (Lombardo, 1987). Gibson's (1950) conception of the visual information provided by such "higher order variables" as a texture density gradient was based on the idea that gradients of repeated structural relations specified the projective transformation of a surface onto an image and also specified an intrinsic scaling of the 3-D space in which the surface texture was homogeneous. The same conception was subsequently expanded (e.g., Gibson, 1957) to emphasize the information provided by the continuous transformations of optical flow produced by moving objects and moving observers. These deformations of the optical images were believed to enable the perception of both the structural invariants and the projective transformations associated with the motions of objects and observers in 3-D space.

The essential ideas underlying this conception of geometry were described by the mathematician Killing (1892)⁴:

Every object covers a space at every time. The space covered by one object cannot simultaneously be covered by another object.

⁴We are grateful to Jan Koenderink for bringing this paper to our attention and to Bernd Rossa for translating the paper from the original German.

Every object can be moved. If an object covers the space of a second object at any time, then the first object can cover any space covered by the second object at any (other) time.

Every space (object) can be partitioned. Each part of a space (object) is again a space. If A is a part of B and B is a part of C, then A is a part of C, where A, B, and C may be either spaces or objects. [p. 128]

These three principles, which are the first of eight principles from which Killing derives a general theory of geometry, provide qualitative criteria for defining the equality or congruence of spaces and objects: Two spaces are congruent if and only if they can be covered by the same object. Two objects are congruent if and only if they can cover the same space. Thus, objects and spaces constitute mutually interdependent relational structures. The metric structure of both may be derived from elementary qualitative properties of *differentiability* and *congruence under motion*. (By definition, "motions" are isometric transformation groups.) (Also see Weyl, 1952, and Guggenheimer, 1963, Sect. 11-2.)

This conception of form and space provides a basis for understanding how visual information about the metric structure and dimensionality of objects and spaces may be gained from "motions" or transformations which bring objects at one position in space into congruence with those at other positions. The metric equality of neighboring spaces successively occupied by the same object and the equality of separate parts of an object which successively occupy the same space may be determined from the motions of objects. Accordingly, the dimensionality of visible spaces and objects need not be restricted to the two coordinate dimensions of the image. Rather, the dimensionality may be associated with the number of parameters needed to bring an object at one location in space into congruence with an object at another location.

In certain special cases the metric structure of a given surface patch may be *locally* determined (up to a scalar) by its moving images, independent of global properties of the retinal image as a whole. If the *trajectory* of the moving patch is also a *surface in space-time* with constant curvature equal to that of the object patch, then of course the metric tensor for this spatio-temporal surface remains constant over the surface. Motion of the object patch from one region of the spatio-temporal surface to another does not change the mapping of the surface onto the retina, and the contravariant tensor coefficients for this projective mapping of the object patch and its trajectory onto the retina vary only as one-parameter functions of time. Accordingly, the perspective parameters for embedding the retinal images of this surface into E^3 also vary as one-parameter functions of time or of retinal position (which are correlated in this case). The simplicity of these relationships between the differential structure of the object surface, its trajectory in space-time, and the retinal images of these surfaces involves sufficiently few unknown perspective parameters that these are determined by the invariance of the metric tensor of the surface patch under motion. That is, suppose that V_0 and P_0 are the Jacobian matrices for the visual and perspective coordinate transformations, respectively, for an initial retinal image of the surface patch, and suppose that V_t and P_t are the corresponding matrices for a second retinal image of the same surface patch following a one-parameter motion onto another position along its constant-curvature trajectory. The equivalence of the geometric structure of the two retinal images can be expressed by the equation

$$V_0^T P_0^* V_0 = V_t^T P_t^* V_t$$

where $P_t = m(P_0)$ is a one-parameter transformation of P_0 . This matrix equation involves four independent linear equations in four unknowns—the three independent perspective parameters of P_0 and the transformation of these by the parameter t .

Specific examples of this special case include a sphere that rotates around an axis (different from the direction of gaze) through its center (e.g., Lappin, Doner, and Kottas, 1980; Doner, Lappin, and Perfetto, 1984) and planar patterns that rotate within the same plane (Lappin and Fuqua, 1983) tilted with respect to the retinal image. In both of these cases the time-varying set of positions of the surface patch form a surface of revolution in space-time generated by a one-parameter transformation group (the magnitude of the rotation). In general, the metric tensor for the images of the moving surface patch remains invariant under the motion (i.e., its Lie derivative is zero) if and only if the vector field of this group of isometries (the "Killing vector") is a one-parameter group that generates a *surface of revolution* (Guggenheimer, 1963, pp. 272-273). Thus, because the moving object forms a surface whose images are generated by a one-parameter transformation, the perspective parameters for embedding this spatio-temporal surface into E^3 are determined up to a scalar by the invariant metric structure of the given surface patch. Indeed, the experimental results of Lappin, Doner, and Kottas (1980) and Doner, Lappin, and Perfetto (1984)—for the perceived shape of a random-dot sphere rotating about a vertical axis through its center—and of Lappin Fuqua (1983) for the perceived inter-point distances among three collinear points rotating in a plane—demonstrated just this invariance of visually perceived metric structure under motion even though the optical displays contained unnaturally exaggerated amounts of polar projection.

In general, however, the metric structure of moving objects cannot be recovered from only local properties of their retinal images. Instead, the perspective parameters of the projection from E^3 onto the retina must be recovered more global constraints on the images. Perspective projection from E^3 onto a plane produces a *hyperbolic geometry* in the plane, where mutually parallel lines converge toward a common vanishing point and all sets of parallel lines converge toward a common horizon line. The position of this horizon line in the visual field is equal to the observer's eye-height. Accordingly, all lines parallel to the observer's motion through the 3-D environment converge toward a common vanishing point on the horizon that specifies the observer's momentary position and trajectory through the visible environment. The images of such parallel lines in E^3 are generated by the retinal image trajectories of features of stationary environmental objects as the observer moves through the environment. Thus, the location of this horizon line and of such vanishing points constitute parameters that characterize the given hyperbolic space and its relation to E^3 . Like Euclidean space, hyperbolic space is also characterized by congruence and isometry of form under motion. Thus, congruence relations among visible forms must specify this global perspective embedding of the retinal image into E^3 . Although we have not yet completed the mathematical analysis of this situation, the following illustrations may help to convey the rationale for this conception of the geometry of vision.

4. CONGRUENCES IN IMAGES

The potential for constructing spaces from congruences among imaged forms has been wonderfully illustrated by M. C. Escher. For example, he has often used translational symmetry of a replicated form to define a 2-D plane. Both the metric structure of this space and also its 3-D orientation parallel to the image plane are specified by the translational symmetry. The elementary

component form is also defined by its recursion in the image rather than by the familiarity of the form itself.

Symmetries in 3-D Euclidean space are exhibited in figure 1, where the congruence of swan-like component forms is obtained by translations and rotations in a 3-D space. The 3-D metric structure of the space is implied by the congruence of the recurring forms in separate regions of the space. The perspective mapping of this space onto the 2-D image plane is also induced by this congruence of the component forms. Thus, the perspective trigonometry is derived from the congruence; the fundamental property is the congruence rather than the trigonometry.

In the preceding example, congruence is defined among stationary and concurrent forms. The "motion" that brings a form in one location into congruence with a form in another location is abstract, rather than an actual trajectory in space-time. If one generalizes the concept of an image from a stationary 2-D spatial array to a space-time volume in which the spatial structures are extended in time, then the same principle of congruence illustrated in Escher's art can be applied to the specification of spaces by the motions of single forms.

The schematic diagram in figure 2 illustrates three conceptually different types of congruence in images. Figure 2B is like that in the Escher print, where the image is a stationary 2-D pattern in which a single cube-like structure is recursively positioned at a sequence of neighboring spatial positions. The 3-dimensionality of the space is induced by the continuous linear change in the 2-D lengths of the contours of the cube as a function of its position in the image plane. This linear relation between 2-D length and position corresponds to a particular perspective mapping of 3-D space onto the image plane. Thus, the continuous linear relation among neighboring regions of the image of a single connected surface specifies the perspective mapping of a 3-D space onto the 2-D image.

In figure 2A the same perspective mapping is defined by a temporal sequence of spatial images as the cube is translated through space from position P_1 to position P_n . The linear transformation that corresponds to the perspective projection of a plane slanted in depth is now specified by a function in space-time, though the geometric relation between the image and the depicted space obviously is essentially the same as in figure 3B. In both cases, relationships among neighboring image regions correspond to relationships among neighboring regions of a smooth surface. The perspective relation between the image and the 3-D space in which the surfaces, objects, and motions reside is specified by the linear relationship between the lengths of the contours and their positions in the image.

Figure 2C illustrates a slightly different case in which the structure of a space is specified by congruences among simultaneous *motions* of separate forms at separate locations in the image, as if the forms were connected and moved in 3-D space. This situation might be produced, for example, by motions of the observer or image plane (e.g., a movie or video camera) within a 3-D environment. In this example two cubes, at positions P_1 and P_n in 3-D space, are simultaneously displaced in a sequence of four successive translations. The perspective mapping from the 3-D space in which these events occur onto the 2-D image of the events may be specified by the functional relation between the magnitudes of the velocities and their locations in the image plane. Although the forms at positions P_1 and P_n in this particular illustration are both cubes that are potentially congruent under the same transformations that would bring the motions of the two cubes into congruence, this spatial congruence is not necessary and provides in this case an addi-

tional redundant specification of the perspective transformation of the 3-D space onto the 2-D image.

The geometric relation between the concurrent motions of just two forms as in figure 2C is not generally sufficient to specify the perspective transformation that has yielded the observed spatio-temporal image. By the fundamental theorem of plane perspectivity (Delone, 1963), the perspective mapping of four points in general position (where no three points are collinear) in one image plane onto a corresponding set of four points in another image plane is necessary and sufficient to ensure that all of the remaining points are in isometric correspondence in the two planes. Thus, for a set of four or more points in a single plane, the concurrent motions of the images of these points in another plane are in principle sufficient to specify the perspective transformation between these two planes and to specify the metric structure of the spatial relations within these planar images.

This geometric relationship endows spatial as well as moving images with considerable capacity for carrying information about the geometric structure of the environmental surfaces depicted in the images: The geometric structure of an infinitesimally small patch on any arbitrarily curved but smooth surface may be locally approximated by a tangent plane at that location, and the perspective mapping of this tangent plane onto an image plane may be described by a linear coordinate transformation. The parameters of this linear transformation vary with the relative 3-D orientation (the direction of tilt and the magnitude of slant) and distance of the environmental surface in relation to the image plane. The perspective parameters which embed the image of the surface into E^3 and thereby determine the metric structure of the surface are those parameters that will yield the self-congruence of the same object at different locations within the depicted scene.

5. EXPERIMENTAL EVIDENCE

In addition to the evidence provided by the illustrations, by everyday visual experience in viewing both natural environments and artificial spatial displays, and by the capabilities of moving observers to coordinate their actions with the identities, positions, and trajectories of environmental objects, the hypothesis that perceived geometric structure derives from the congruences of moving and movable objects is also supported by experimental evidence. A vast amount of experimental evidence appears consistent with this hypothesis, but we mention here only a few experiments that seem to provide more direct support for this hypothesis.

One of the relevant investigations is that of Cutting (1987). Judgments of the apparent rigidity of rotating rectangular solids were evaluated in a variety of experimental display conditions, including both rigidly and nonrigidly rotated figures and displays that simulated varying degrees of polar versus parallel projection, and varying degrees of slant of the projection screen relative to the direction of the perspective convergence point. He found good discrimination of rigid versus non-rigid figures in displays with approximately parallel projection, essentially independent of the degree of simulated screen slant (90° , 67° , 45° , or varying between 80° and 55°), even when the simulated slant was varied sinusoidally during a given trial. Although the figures appeared to move nonrigidly in conditions with polar projection onto screens slanted at 45° , the results generally demonstrated the robustness of perceived structural rigidity under at least moderate screen slants and moderate viewing distances. These results challenge many conventional assumptions about the geometrical information for perceiving the spatial structure of form. Cutting concludes

that these results probably reflect the insensitivity of vision to the distortions produced by optical projections, but this interpretation rests upon assumptions about the definition of visual space by the metric structure of 2-D display screens and retinae. An alternative interpretation is that vision is very sensitive to spatial relations defined in another way—by the congruence of form under perspective transformations.

Evidence that vision is indeed very sensitive to the spatial structure of moving forms and that this structure is associated with invariant spatial relations in depth rather than the projected 2-D positions is provided by experiments reported previously by Lappin and Fuqua (1983). They evaluated observers' acuities in detecting a displacement (a stationary offset in 3-D space) of a point from the 3-D center of an imaginary line segment defined by moving patterns of three collinear points. The points were rotated in computer-controlled CRT displays as if around an axis slanted in depth by amounts varying between trials from 0° (no slant) to 60°. Very small displacements were accurately detected—displacements greater than 1% of the 3D distance between the two outer points could be detected above chance, and displacements of 4% were detected at approximately 90% accuracy. The essential 3-dimensionality of the perceived spatial relations was demonstrated by the following findings: (1) Detection accuracy was independent of either the magnitude or variability of the slant of the axis of rotation in depth. (2) Distance-like measures of the detection accuracy (similar to the signal detectability measure d') were linearly related to the physical distance of the displacement in 3-D space, with discriminability being proportional to physical displacement distances above about 1%. (3) The accuracy for detecting any given displacement was the same in displays with parallel and with polar perspective, although in the latter displays points centered in 3-D depth were not centered in the projected 2-D images. The differences in spatial positions between the parallel and polar displays were visually resolvable, however. (4) When the task required detection of displacements from the projected 2-D centers of the line segments in displays with polar projections, accuracies were not significantly above chance. The subjective appearance of the latter displays was that the three points were still seen as rotating in depth, but the middle point appeared neither centered nor rigidly attached to the two outer points.

Thus, these findings suggest that vision may often be unaffected by the 2-D optical "distortions" in cinema not merely because these spatial differences cannot be resolved by vision, but because they do not constitute the geometrical information for perceiving the spatial structure of moving patterns. Apparently, perceived spatial structure derives from congruences of form under perspective transformations.

Evidence about the role of such congruences in *stereoscopic* form perception has been provided by recent experiments described by Lappin (in preparation). The purpose of these experiments was to determine whether the stereoscopic perception of 3-D structure might be shaped by the congruences of form associated with motion in depth, rather than by the binocular disparities as such. The experiments were motivated by the theoretically challenging fact that for any given magnitude of binocular disparity between the horizontal separations of a pair of points in each eye, the associated separation in depth increases rapidly and nonlinearly with the viewing distance from the observer to the points in question: How then is the stereoscopic perception of form and depth calibrated for variations in viewing distance? Does this require "interpretations" of retinal disparities based on extra-retinal information about the viewing distance? Alternatively, might the perceived geometric structure of surfaces in depth be based on the invariance of the intrinsic geometric structure of the surface under the perspective transformations associated with stereoscopic disparities and with motions in depth? The theoretical problem is related to those in understanding the apparent "paradoxes" of cinema.

In one of these experiments, observers were presented with two very slightly different ellipses, in which the vertical axis was either 3% greater or less than the length of the horizontal axis. These ellipses were displayed as if in a plane slanted in depth by either 50° or 60° varying randomly from one trial to the next. Thus, the projected forms were always elliptical, depending on the magnitude of the slant as well as on the shape of the ellipse as measured in its own plane in depth. Stereoscopic information about the shapes and slants of these patterns was also manipulated by random variations in the magnitude of the disparities with which the forms were displayed, using disparities that were appropriate for either one-half or one-quarter of the actual viewing distance at which the patterns were seen. Thus, there were eight alternative stimulus patterns which randomly varied between trials.

There were four main experimental conditions—in which the forms were either rotated in depth or were stationary, and in which the experimental task was either *shape-discrimination* between the two alternative ellipses or *disparity-discrimination* between the two alternative disparity values. If stereoscopic information about 3-D structure is scaled by the congruences of moving forms, then shape discrimination should be accurate when the forms were rotated in depth, independently of the distortions and variability produced by the exaggerated binocular disparities. Indeed, this is just what happened: Shape discriminations were very accurate when the forms were moving, and were uncorrelated with the variations in either slant or disparity. Not surprisingly, shape discriminations were near chance accuracy when the forms were stationary because of the perceptually inseparable conjoint effects of variations in slant and disparity. For the disparity-discrimination task, however, motion had the opposite effects: Discriminations between the two alternative disparity values were more accurate for the stationary than for the moving forms, evidently because the congruence of the moving forms tended to obscure differences between the stationary disparity spaces.

Thus, these results indicate that the visual scaling of 3-D structure from stereoscopic disparity derives from the congruences of the perspectively changing forms. Analogous to the case for stationary pictures and optic flow patterns, binocular disparity per se may have only an indirect relation to the perceived depths.

REFERENCES

- Cutting, J. E. (1987) Rigidity in cinema seen from the front row, side aisle. J. Exp. Psychol: Human Percept. Perform. **13**, 323-334.
- Delone, B. N. (1963) Analytic geometry. In A. D. Aleksandrov, A. N. Kolmogorov, and M. A. Lavrent'ev (Eds.) Mathematics: Its content, methods, and meaning. Cambridge: MIT Press.
- Doner, J., Lappin, J. S., and Perfetto, G. (1984) Detection of three-dimensional structure in moving optical patterns. J. Exp. Psychol: Human Percept. Perform., **10**, 1-11.
- Einstein, A., Lorentz, H. A., Weyl, H., and Minkowski, H. (1923/1952) The principle of relativity. New York: Dover.
- Gibson, J. J. (1950) The perception of the visual world. Boston: Houghton-Mifflin.
- Gibson, J. J. (1957) Optical motions and transformations as stimuli for visual perception. Psychol. Rev., **64**, 288-295.
- Guggenheimer, H. W. (1963) Differential geometry. New York: McGraw-Hill.
- Killing, W. (1892) Ueber die Grundlagen der Geometrie. J. für die reine und angewandte Mathematik, **109**, 121-186.
- Koenderink, J. J. and van Doorn, A. J. (1975) Invariant properties of the motion parallax field due to the movement of rigid bodies relative to the observer. Opt. Acta, **22**, 773-791.
- Koenderink, J. J. and van Doorn, A. J. (1976a) Local structure of movement parallax of the plane. J. Opt. Soc. Amer., **66**, 717-723.
- Koenderink, J. J. and van Doorn, A. J. (1976b) Geometry of binocular vision and a model for stereopsis. Biol. Cyber., **21**, 29-35.
- Koenderink, J. J. and van Doorn, A. J. (1976c) The singularities of the visual mapping. Biol. Cyber., **24**, 51-59.
- Koenderink, J. J. and van Doorn, A. J. (1977) How an ambulant observer can construct a model of the environment from the geometrical structure of the visual inflow. In G. Hauske and E. Butenandt (Eds.) Kybernetik. Munich: Oldenberg.
- Koenderink, J. J. and van Doorn, A. J. (1980) Photometric invariants related to solid shape. Opt. Acta, **27**, 981-996.
- Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A. (1971) Foundations of measurement. New York: Academic Press.

- Lappin, J. S. (in press) Perceiving the metric structure of environmental objects from motion and stereopsis. In R. Warren and A. H. Wertheim (Eds.) The perception and control of self-motion. Hillsdale, N.J.: Lawrence Erlbaum.
- Lappin, J. S. (in preparation) The scaling of stereoscopic form by congruence under motion.
- Lappin, J. S., Doner, J. F., and Kottas, B. L. (1980) Minimal conditions for the visual detection of structure and motion in three dimensions. Sciences, 209, 717-719.
- Lappin, J. S. and Fuqua, M. A. (1983) Accurate visual measurement of three-dimensional moving patterns. Science, 221, 480-482.
- Lombardo, T. J. (1987) The reciprocity of perceiver and environment: The evolution of James J. Gibson's ecological psychology. Hillsdale, N.J.: Lawrence Erlbaum.
- Luce, R. D. (1978) Dimensionally invariant numerical laws correspond to meaningful numerical relations. Phil. Sci., 45, 1-16.
- Luce, R. D. and Narens, L. (1983) Symmetry, scale types, and generalizations of classical physical measurements. J. Math. Psychol., 27, 44-85.
- Misner, C. W., Thorne, K. S., and Wheeler, J. A. (1973) Gravitation. San Francisco: Freeman.
- Weyl, H. (1952) Space-time-matter. New York: Dover.

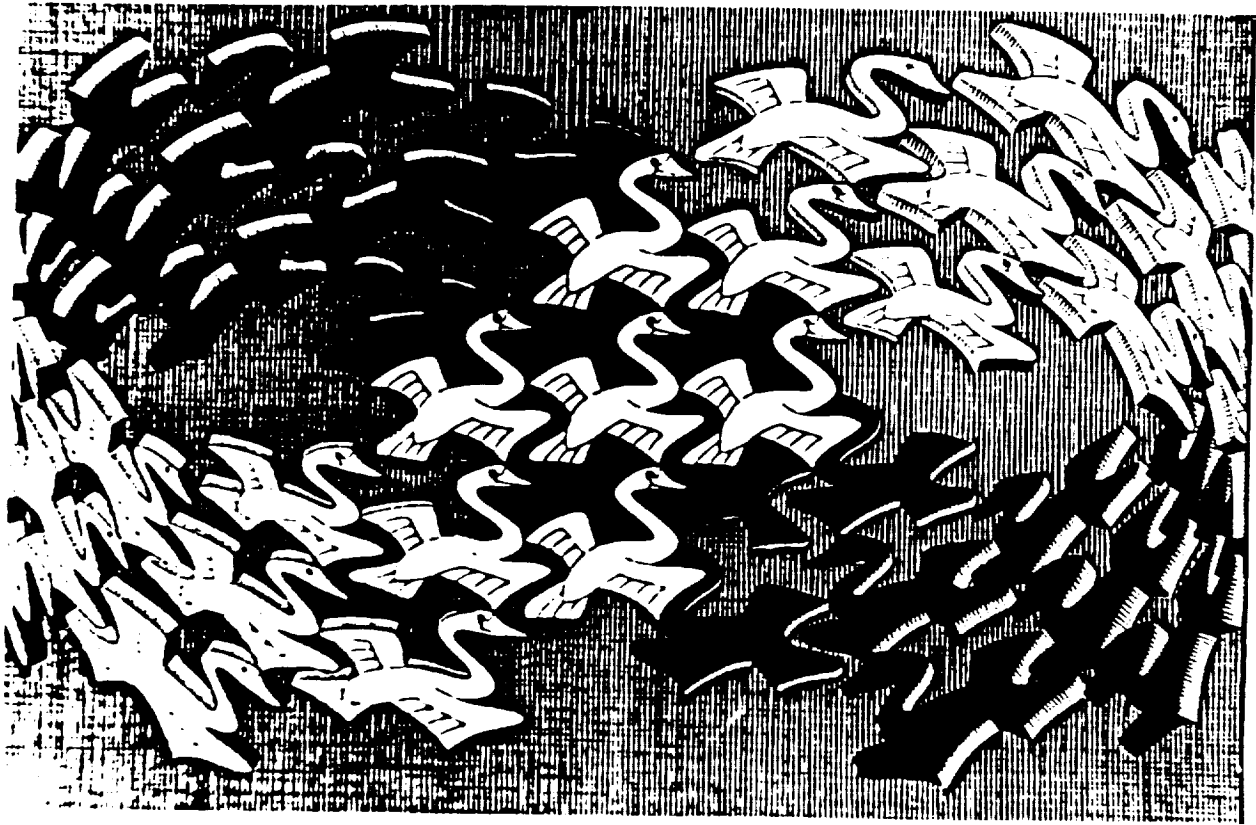


Figure 1.— "Swans," etching by M. C. Escher, 1956. © 1988, M. C. Escher heirs/Cordon Art-Baarn-Holland.

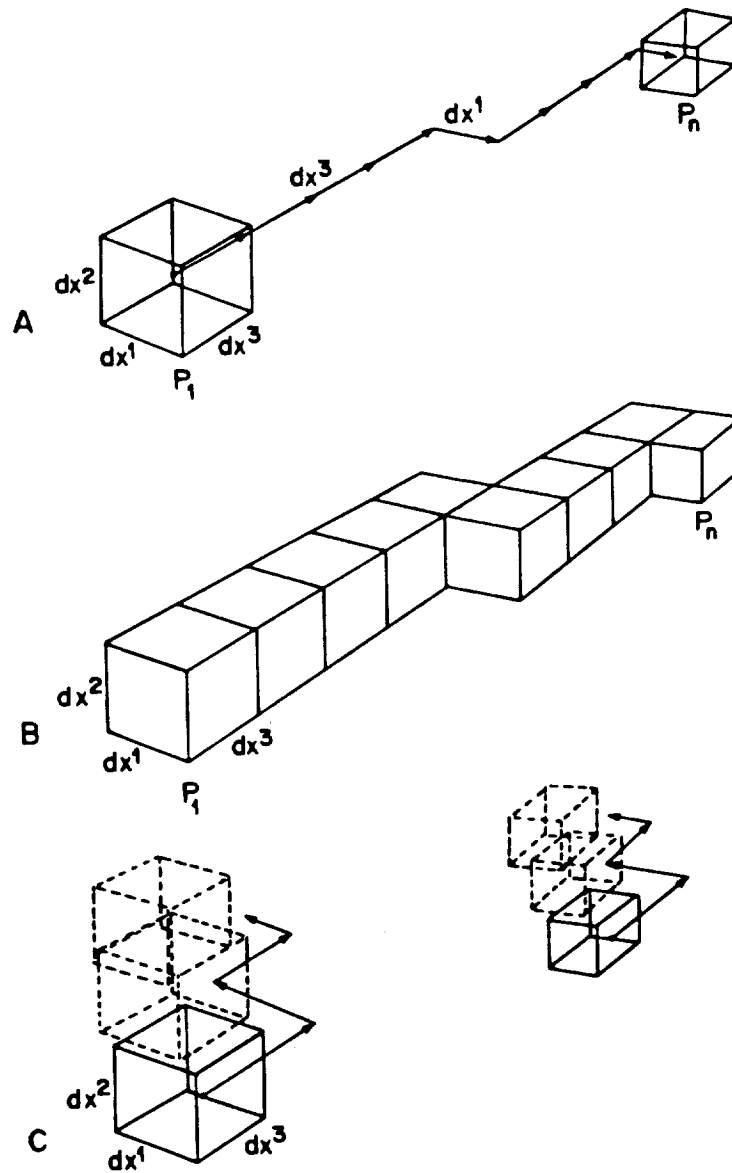


Figure 2.— Three types of congruences in images. (A) A cube in position P_1 is moved in a temporal sequence of displacements through 3-D space to position P_n . A single object appears in a trajectory through space-time. (B) The same cubic form as in A appears simultaneously in positions P_1 and P_n , connected in this case by a spatial series of cubes. A 3-D space is defined by the congruences of the spatial series of repeated component forms. (C) Two objects are moved concurrently by a sequence of displacements as if rigidly connected. The 3-D structure of the space is indicated in this case by the congruence of the motions in the separate spatial regions rather than by the congruences of the spatial forms as in the other two panels.

