Symmetry Detection and the Perceived Orientation of Simple Plane Polygons

Paul Kube

Institute of Cognitive Studies
and the
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

1. Introduction

Recent advances in the computational theory of vision have been impressive. However, some simple and basic aspects of human visual performance have yet to be explained. For example, the details of human performance in seeing slanted isolated textureless plane polygons *as* slanted plane polygons are not predicted by any current theory. Below, five constraints on any correct theory in this domain are extracted from psychological experiment and contrasted with predictions of three of the best current computational theories; no theory meets them all. A detector model is proposed which can qualitatively account for the evidence. The model is one obtained by simple elaboration of a symmetry detection model introduced by Palmer [15] to account for a number of other perceptual phenomena in a unified way.

2. What experiment says

The study of the perception of the perspective projection of simple polygons dates from before the classic monograph of Stavrianos [16]. By 1966 the literature was fairly large (see, for example, Freeman's bibliography [10]), and continued to grow into the early seventies. Most studies in this tradition were restricted to stimuli perceptually indistinguishable from rectangles centered in the line of sight and rotated on an axis parallel to one side; i.e., stimuli which would project under perspective into the frontal plane as isosceles trapezoids. Data from the presentation of other kinds of contour stimuli is virtually absent. However, it appears that facts about the perception of even such simple objects are hard to account for in a simple theory.

Data from this body of work disagree on some points (concerning, for example, how tight the coupling is between a figure's perceived shape and perceived slant, or how the accuracy of slant judgments varies with stimulus size), but there are at least two unequivocal results worth noting. As summarized by Flock et al., "the single untextured rectangular shape when viewed monocularly without parallax constitutes too great a degree of impoverishment to elicit accurate slant judgments from the human visual system. . ." [9 p. 58] That is, human viewers aren't very good (mean regression no better than about 0.7 in the reported experiments) at seeing the precise slant of slanted rectangles when outline is the only cue. On the other hand, they are fairly likely to see slanted rectangles,or figures which project outlines indistinguishable from slanted rectangles, *as slanted*; see [5]. These simple facts impose two constraints (C1 and C2, below) which must be met by any candidate theory of human vision.

Two exceptions to the isosceles-trapezoid-stimulus tradition yield three additional constraints (C3, C4, and C5 below). An experiment reported by Attneave [1] featured the solicitation of slant judgments from subjects upon viewing frontal-plane parallelograms of various shapes and orientations. Rectangles at most orientations and aspect ratios showed great resistance to being seen as slanted in depth at all. Highly nonrectangular parallelograms were relatively easy to see as slanted figures. In a pilot study at the Berkeley ICS, subjects are shown outline quadrilaterals like those in Figure 1.

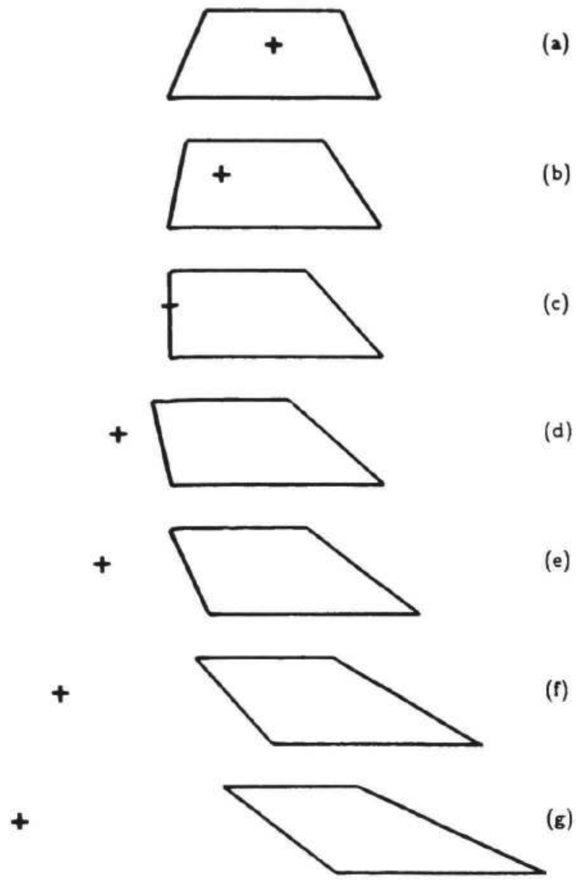

Figure 1, a-g

A series of seven perspective projections of a square in a plane with slant $\sigma = 60$ and tilt $\tau = 90$ degrees. The +'s are not part of the stimulus; they merely mark here the origin of the image plane for each projection. When a figure is foveated with the center of the lens above the origin, and such that the figure subtends fifteen degrees of arc vertically, the retinal image approximates viewing the outline of a meter-square tabletop, with near edge parallel to the image plane, at a distance of about two meters. (Of course, these reproductions need to be scaled before being used as experimental stimuli; at the size reproduced here, the focus is only about 4 cm. above the page.)

One of the figures at a time is monocularly presented to a subject so that the contour falling on the retina is the projection of a square slanted with respect to the frontal plane; the subject is asked to report whether or not she sees the figure as flat and slanted (i.e., lying in a plane not parallel to the frontal plane). Preliminary results are shown in Figure 2, with the vertical axis indicating increasing ease in seeing the figure as a slanted plane polygon. Some projections are considerably easier to see in this way than others, though all are equally *correct* projections of slanted squares.

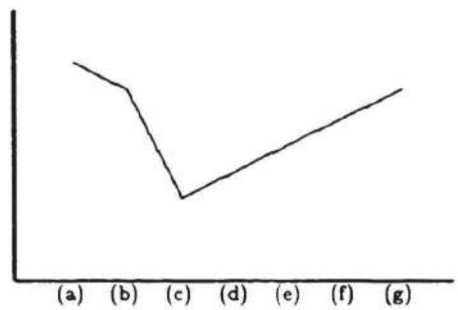

Figure 2

Relative "goodness of depth illusion" for the seven stimuli shown in Figure 1. Goodness of illusion increases upward along the vertical axis.

Thus, from experimental results, there are at least these five straightforward constraints that any theory of human visual capacity must meet:

C1. Viewers are fairly bad at judging the precise slant of slanted rectangles, even when they know they are viewing slanted rectangles.

C2. Figures which project isosceles trapezoids into the frontal plane can be easily seen as plane figures slanted in depth.

C3. Figures which project rectangles into the frontal plane are very hard to see as slanted in depth.

C4. Figures which project highly nonrectangular parallelograms into the frontal plane are easy to see as slanted plane figures.

C5. **Of the actual perspective projections of a slanted rectangle at various orientations, some are more difficult to see as slanted plane figures than others.**

3. What theory says

Following years of concentrated work in the development of polyhedral scene understanding systems (for example [11] [8] [14]) which had nothing to say about the perception of the orientation of isolated surfaces, recently there has been a flowering of interest in the construction of computational theories which demonstrate how, under plausible assumptions, depth information can be extracted from retinal images, and which sometimes do have consequences for the perception of slant of simple isolated polygons. To their credit, these theories typically aim for considerable generality and have these consequences, if they do have them, among many others of interest. The ones whose predictions will be examined below with respect to the experimental lessons of the previous section are all expressions of one sort or another of kind of principle of *Prägnanz*: they predict that an image will be interpreted as having been projected by the object which, of all possible projecting objects, best meets some criterion of simplicity.[1]

Several recent theories of this sort are not open for consideration because they are not even well-defined for polygonal stimuli. Witkin [19], for example, proposes that the chosen interpretation of a projection should be the one that maximizes the uniformity of distribution of orientations of contours in the scene; however, his mathematics require that the contour orientations be statistically independent, which is certainly not the case for simple polygons. Barrow and Tenenbaum [4] suggest that the contour which minimizes the integral of the square of its curvature over its length should be chosen as the interpretation; but curvature is either infinite or undefined at each vertex of a polygon, and zero everywhere else, so it fails to distinguish between any alternative projections. Barnard and Pentland [3] propose that the selected interpretation be one composed of circular arcs, and they offer an ellipse-fitting algorithm that achieves such an interpretation in some cases; but it fails on polygons. Three theories which do make predictions for the perception of simple polygons are considered below.

[1] Or, at least, some criterion that is correlated with simplicity. Steve Palmer has suggested to me that, for example, the area-divided-by-squared-perimeter measure of Brady and Yuille (see below) should not be seen as a simplicity measure itself, but rather as a measure the maximization of which happens, in a range of cases, to pick out simple figures.

3.1. Kanade

A bilaterally symmetric figure is one whose shape is invariant over reflection about a line, the axis of bilateral symmetry. Consequently, such a figure is composed of pairs of points which lie at equal distances in opposite directions from the symmetry axis on lines that all meet the axis of symmetry at a 90 degree angle. If a figure is bilaterally symmetric, any affine transformation of the figure will be *skew-symmetric*: paired points in a skew-symmetric figure lie on lines (the *skew-transverse axes*) which all meet the skew-symmetry axis at some arbitrary angle, not necessarily 90 degrees. Since orthographic projection is an affine transformation, bilateral symmetries in the scene will become, under orthographic projection, skewed symmetries in the image. This fact is exploited by Kanade [12] [13], to derive constraints on perceived object orientation upon detection of image skewed symmetry, under the *Prägnanz*-like assumption that the perceptual system should prefer to see bilateral symmetries whenever possible.

To see how this constraint on orientation from skewed symmetry is supposed to work, consider the polygon in Figure 3 with its indicated skew symmetry. In this case, either dashed line can be taken as the skewed-symmetric axis, and the other as a skewed-transverse axis. The Kanade

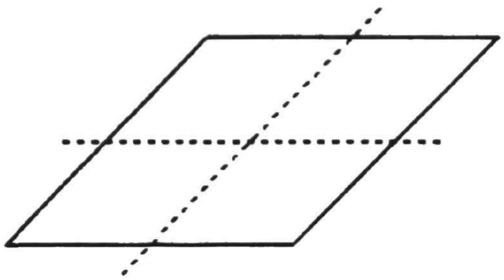

Figure 3

A skew-symmetric polygon. Dashed lines show axes of skewed symmetry.

assumption is that a skewed symmetry observed in an image was projected by a real symmetry in the imaged scene, and so the axes of skewed symmetry in the image are projections of the axes of bilateral symmetry in the scene. But axes of bilateral symmetry meet at right angles; so the assumption constrains the orientation of the plane containing the bilateral symmetry to be such that lines lying in it and meeting at right angles could have orthographically projected the observed axes of skewed symmetry in the image.

Kanade has shown that this is not sufficient to fix a unique interpretation for the orientation of the viewed figure; instead, the angle formed by the skewed-symmetric and skewed-transverse axes determines an orientation which must lie on a hyperbola (or, if the angle is 90 degrees and so the image is already bilaterally symmetric, a pair of perpendicular lines) in gradient space. (Here, gradient space is a two-dimensional space in which each point represents the orientation of a plane, and each of a set of parallel planes maps to the same point. The mapping is standardly defined as follows. Let there be three-dimensional cartesian coordinates in space such that the image plane (e.g., the 'plane' of the retina) satisfies $z=0$. Now an arbitrary plane in space will satisfy an equation of the form $px+qy+d=z$, for some value of p, q, and d; and the equation of any plane parallel to this one will differ only in the value of d, so specifying p and q suffice to specify the orientation of the plane. The point (p,q) is then the map of the plane in gradient space. Alternatively, the arctangent of the length of the vector $<p,q>$ gives the slant σ of the plane, and the angle of the vector clockwise from the p-axis gives the tilt τ; that is, to talk of orientation in terms of slant and tilt is just to impose polar coordinates on gradient space.) In the present case, the skewed-symmetry axes shown in Figure 3 constrain the object plane to have an orientation lying on the hyperbola in Figure 4.

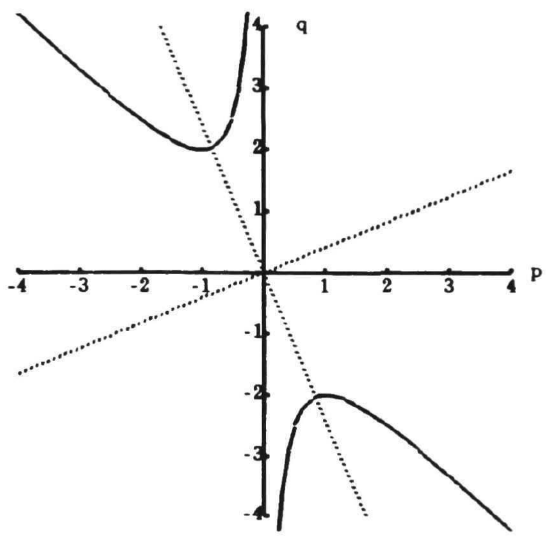

Figure 4

Gradient space constraints on the plane containing the skewed symmetry shown in Figure 3 (solid hyperbolic curve) and in Figure 5 (dotted orthogonal lines).

Thus Kanade would predict that because of the detection of the indicated skew symmetry, Figure 3 will be seen as a rectangle (a parallelogram contralateral bisectors meet at right angles) whose orientation lies on the hyperbola in Figure 4; and Kanade [12] suggests further that, in the absence of further constraints, the perceptual system should select the least slanted of the orientations that are possibile given a symmetry constraint, i. e., an orientation on one of the apexes of the hyperbola: in this case, a slant of about 65 degrees in the direction of one or the other of the Necker-reversal tilts.

This result seems plausible, but there are inadequacies with the approach. Note that the parallelogram of Figure 3 has another pair of skewed symmetry axes, as shown by the dashed lines in Figure 5.

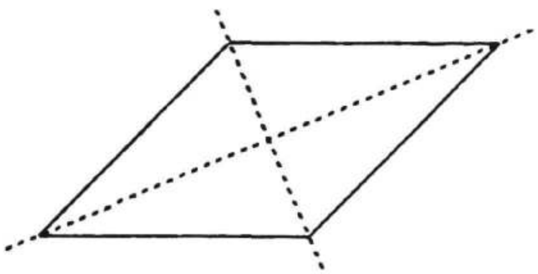

Figure 5

Another skewed symmetry for the polygon in Figure 3.

These happen to be axes of true bilateral symmetry, so they require that the plane of the projecting object have an orientation that falls on one of a pair of perpendicular lines in gradient space, which, for this example, are the broken lines of Figure 4. The theory does not explain how the constraints from these two skewed symmetries should be combined, and no reasonable combination seems to be consistent with all of the experimental constraints:

If minimization of slant is to count more heavily than reconstructing all possible symmetries, then the angle bisector axes of Figure 5 should be the ones that the Kanade constraints are meant to apply to; this gives the prediction of seeing the figure in the frontal plane, i.e., at zero slant. But this would violate constraint C4.

Perhaps, instead, the theory should be understood as predicting that all possible symmetries will be reconstructed, and that as a result the figure will be seen at an orientation that simultaneously satisfies the constraints imposed by both pairs of axes. This would correspond to the intersection, in the p,q plane, of the curves given by the two skewed symmetries; see Figure 4. In this example, this gives the same prediction as minimizing over the range of slants permitted by the contralateral bisector symmetry alone, which was a reasonable prediction.

But this suggestion fails on other examples. *Any* parallelogram in the image plane has two pairs of skewed symmetry axes, one pair connecting the midpoints of its sides and the other connecting its vertices. Now the skewed-symmetry assumption is that a skewed symmetry in the image was projected by a true bilateral symmetry in the scene; and so the skewed-symmetry axes must have been projected by perpendicular axes. But taking the intersection of the constraints imposed by both pairs of axes, this implies that the figure projecting a paralellogram must be a convex quadrilateral whose diagonals, and whose opposite-side bisectors, meet at right angles: i.e., the projecting figure must, in either case, be a *square*.

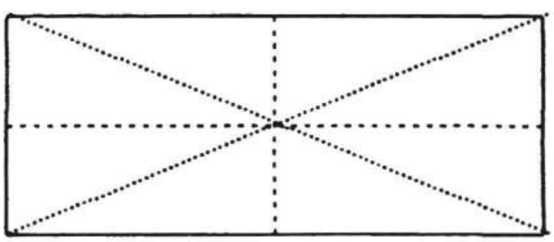

Figure 6

This rectangle, with its two skewed symmetries as shown, has the same slant constraints as the parallelogram in Figure 3

So, in avoiding a violation of C4, the Kanade assumption seems forced to predict seeing *every* nonsquare parallelogram as a square at some nonzero slant. This violates constraint C3, as can be seen by considering Figure 6, a figure whose two pairs of skewed axes are those just those of Figure 3 rotated in the frontal plane. As a result, it has orientation constraints identical to those shown in Figure 4 up to a rotation in the p,q plane, and so the same slant constraints; but it is not naturally perceived as a slanted square in the manner of Figure 3, but rather as a rectangle in the frontal plane.

If instead (and, so far as I can see, without motivation from the theory) opposite-side bisector symmetries are taken as imposing the important orientation constraints, Figures 3 and 6 no longer provide counterexamples. However, the theory would still violate experimental constraint C2. Consider the projected isosceles trapezoid in Figure 1.1; it has only one pair of skew symmetry axes, which are also bilateral symmetry axes, and so the figure should be seen only in the frontal plane. There is nothing in Kanade's theory to account for its being easy and natural to see as a slanted rectangle.

3.2. Brady and Yuille

Brady and Yuille [6] have proposed that an image plane polygon be interpreted as having been projected by the object which, of all possible projecting objects, maximizes the ratio of area to perimeter squared. This measure tends to favor compact, nonelongated figures, and is maximized by squares within the class of quadrilaterals, so their theory entails the perception of a square whenever projectively possible. But, since their mathematics is developed under the assumption of orthographic projection, this (as for Kanade) leads to a violation of constraint C3 when applied to Figure 6. Also, it is easy to show that (under orthography) there is a family of image-plane isosceles trapezoids which are self-maximal over this measure; for these trapezoids, the preferred projecting figure is in the frontal plane, violating C2. Weiss, in a recent paper [18], proposes an improvement on the Brady and Yuille measure, but it also assumes orthography, and it violates C3.

Although the mathematics become more difficult, it is conceptually simple to consider Brady and Yuille measure maximization under perspective, instead of orthographic, projection; then it would predict that all image plane figures which are perspective projections of squares will be seen as squares. But this falls subject to the same criticism we suggest for Barnard's approach in the next section—that is, it would violate constraints C1 and C5.

3.3. Barnard

Barnard [2] suggests that angles in an image plane quadrilateral be interpreted as right angles whenever possible, and shows a way to obtain this interpretation under perspective projection. The problem of this approach, roughly, is that it works *too* well; his Figure 12a (reproduced here as Figure 7) is one that subjects find difficult to see as a slanted plane figure, whereas Barnard's method flawlessly extracts the projectively dictated interpretation of it as a slanted rectangle. The mathematics is impressive, but as a candidate psychological theory it violates both constraints C1 and C5.

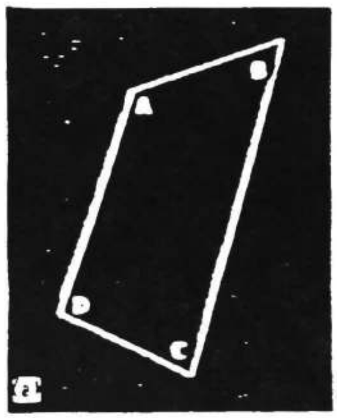

Figure 7

A perspective projection of a rectangle, from Barnard [2]

4. A proposed detector model

If the points made in the previous sections are correct, no single, simple *Prägnanz*-like account of goodness of interpretation of polygonal stimuli suggested in the literature can account for even a very modest range of results from psychological experiment. Some reasons why this might be so are discussed in the final section of the paper. In this section, a system of detectors is described which would exhibit performance consistent with constraints C1 - C5.

Note that C1 - C5 can reasonably claim to all be satisfied by a theory which at least qualitatively predicts the difficulty distribution (graphed in Figure 2) for the stimulus sequence of Figure 1, since this sequence contains both isosceles-trapezoidal (Figure 1.a) and (approximately) skew symmetric (Figure 1.g) stimuli,[2] as long as it is able to independently meet C3. This suggests a detector architecture in which at least some skewed symmetry and perspective gradients – the depth cues which seem to be presented by nonrectangular parallelograms and isosceles trapezoids, respectively – are detected independently and combined to yield a judgment of slant. Palmer [15] has argued that a wide range of phenomena in the psychology of human vision can be accounted for by appeal to a processing model which detects symmetries over members of the Euclidean similarity group exhibited in the stimulus array, certain relations among these symmetries, certain further relations among these relations, and so on. The model sketched here can be seen as an extension of Palmer's Euclidean symmetry detection model to incorporate higher-order detectors for some skewed symmetries and perspective gradients.[3]

[2] Figure 1.g is even closer to a true skew-symmetrical retinal stimulus than it may appear in the reproduction, since with the center of the lens at the focal point on the image plane z-axis and the figure foveated, the retina is slanted with respect to the image plane by about 45 degrees.

[3] Compare Clark, et al.'s [7] "retinal gradient of outline".

The present model supposes three layers of detectors. The lowest level is an array of first-order analyzers (in the sense of Palmer [15]), whose patterns of sensitivity to features of retinal stimulation can be related to one another by transformations from the Euclidean similarity group. For example, each can be taken as detecting, for some location of interest on the retina, the presence of line segments of a certain restricted range of lengths and orientations. (Obviously, such analyzers are interrelatable by translation, rotation, and dilation transformations.)

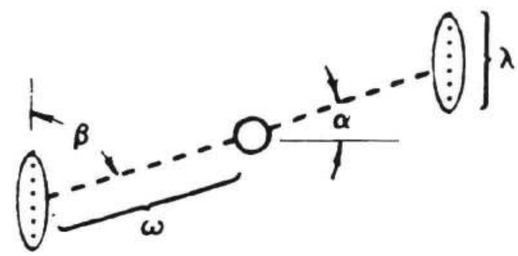

Figure 8

A translational symmetry detector element (TDE).

The second level consists of arrays of two kinds of detector elements, each of which is constructed by simple interconnections between two first-order analyzers. A translational symmetry detector element (TDE) is shown in Figure 8. Such a detector is parameterized by the retina-relative location of its center (represented by the open circle), the length and orientation (λ and β, respectively) of its constituent first-order analyzers (with receptive fields represented by the ellipses), and its own width and orientation (ω and α, respectively). A reflectional symmetry detector element (RDE), shown in Figure 9, is defined by the same set of parameters as a TDE, but its constituent first-order analyzers are related to each other differently: viz. by a reflection about a line through its center and perpendicular to its orientation, instead of a translation.

At the third level, TDE's and RDE's are connected to construct skewed symmetry and perspective gradient detectors – SSD's and PGD's, respectively. A

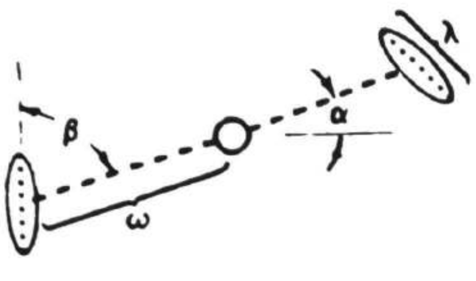

Figure 9

A reflectional symmetry detector element (RDE).

constant-width SSD is shown in Figure 10. Its constituent TDE's differ from each other only by a translation parallel to the favored direction of their constituent first-order analyzers. A bilaterally symmetric PGD, as depicted in Figure 11, is composed of RDE's which differ only by a composition of an appropriately related translation and dilation.

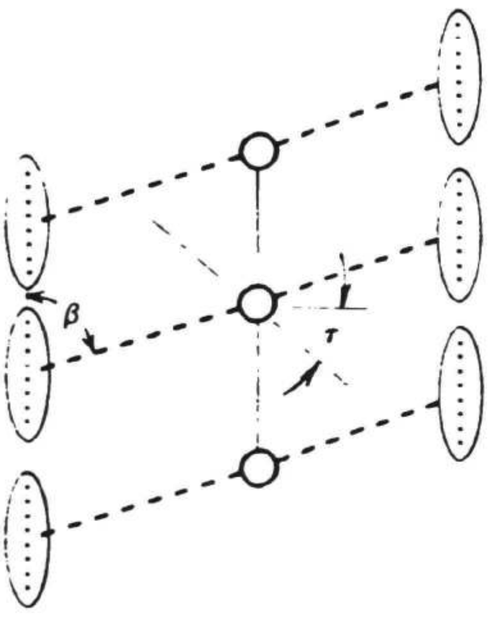

Figure 10

A constant-width skewed symmetry detector (SSD).

A detector d has associated with it a level of activation $A(d)$ which depends on the activations of its constituent detectors (or, in the case of first-order analyzers, directly on patterns of sensory stimulation). Let the activation of a TDE or RDE d be a suitable nonlinear function of the activation of its associated first-order analyzers f_1, f_2:

$$A(d) = \begin{cases} A(f_1)+A(f_2) & \text{if } A(f_1)>\theta \text{ and } A(f_2)>\theta \\ 0 & \text{if } A(f_1)<\theta \text{ and } A(f_2)<\theta \\ -\left|A(f_1)-A(f_2)\right| & \text{otherwise,} \end{cases}$$

where θ is a suitable threshold value. (What's important is that evidence for, evidence against, and the absence of evidence for or against the existence of a symmetry each be encodable in a second-level detector's activity.) Let the activation of a SSD or PGD be a monotonic function (say, the sum) of activations of its constituent detector elements.

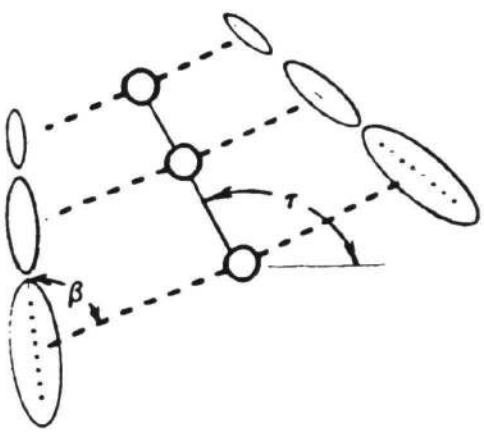

Figure 11

A bilaterally symmetric perspective gradient detector (PGD).

Now the patterns of activations of SSD's and PGD's can give rise to orientation judgments consistent with constraints C1 - C5 as follows. Suppose the SSD or PGD with the highest activation level, if it exceeds some threshold, is taken as encoding the viewed figure's most likely orientation in depth.[4] The tilt direction suggested by that detector is then just τ in Figures 10 or 11: the bisector of the angle between the SSD's axis and the favored direction of its constituent TDE's first order analyzers, or the axis of the PGD. Slant is a not such a simple function of detector parameters; though it is, under reasonable assumptions, monotonic increasing from zero as β in the Figures moves away from 90 degrees for both kinds of detectors. (This is suggestive of Stevens' [17] findings indicating that tilt judgments are more accurate than slant judgments.) This gives the desired results: C1 is explained, since slant is only imprecisely correlated with detector activation; C2 and C4 are satisfied, since PGD's and SSD's are suitably activated by isosceles trapezoids and nonrectangular parallelograms in the retinal image, respectively; C3 is met, since the most active detectors under stimulation by frontal plane rectangles will be a SSD or PGD with $\beta = 90$ degrees; and, since for projections intermediate between the shapes favored by the two types of detectors there may be no detector very activated, C5 is accounted for.

5. Discussion

We have extracted from the reports of psychological experiments five uncontroversial facts about human performance in the perception of simple polygons slanted in depth, and argued that no current theory of orientation from contour is consistent with all of them. We have proposed a detector model which would exhibit performance consistent with the constraints.

The model, however, is not a computational theory; it is a process model which in fact depends for its intelligibility on computational theory (without Kanade's theorems about skewed symmetry and facts about perspective projection, the correlation between β in Figures 10 and 11 and detected slant would be unexplained). It succeeds where the reported theories fail simply because it embodies an interaction between distinct kinds of evidence for orientation in depth. But this success says something about vision: the generation of orientation judgments of simple figures is not simple. The computational theories discussed in this paper are, individually, inadequate; but also the elementary two-factor model proposed here is grossly inadequate as a general account of vision, or even as an account of the monocular perception of simple plane figures. It fails on nonquadrilateral polygons and ellipses; it says nothing about orientation effects (rectangles, for example, present somewhat better depth illusions when their sides are not aligned with environmental horizontal and vertical); it's silent about how complexes of polygons might be perceived as three dimensional solids, and even about how an isolated polygon can be seen at different orientations in depth at different times. But it's unlikely that all the phenomena of monocular vision are going to be subsumed by a single computational principle. What's needed is a theory of what underlies each of them, and a theory of their interaction.

[4] Of course, the detectors need to be thickly distributed enough to respond to stimulus figures at various image locations, orientations, and scales.

Acknowledgements

The research reported here was supported by a grant from the Alfred P. Sloan Foundation to the Institute of Cognitive Studies at Berkeley. I want to thank Steve Palmer for comments on a draft of this paper, and for stimulating discussions.

REFERENCES

1. Attneave, F., Representation of Physical Space, in *Coding Processes in*

Human Memory, A. W. Melton and E. Martin (editor), V. H. Winston and Sons, Washington, D. C., 1972.

2. Barnard, S. T., Interpreting Perspective Images, *Artificial Intelligence 21* (1983), 435-462.

3. Barnard, S. T. and A. P. Pentland, Three-dimensional Shape from Line Drawings, *Proceedings, International Joint Conference on Artificial Intelligence*, 1983.

4. Barrow, H. G. and J. M. Tenenbaum, Interpreting Line Drawings as Three-Dimensional Surfaces, *Artificial Intelligence 17* (1981), 75-116.

5. Beck, J. and J. J. Gibson, The Relation of Apparent Shape to Apparent Slant in the Perception of Objects, *Journal of Experimental Psychology 50* (1955), 125-133.

6. Brady, M. and A. Yuille, An Extremum Principle for Shape From Contour, A.I. Memo No. 711, M.I.T., 1983.

7. Clark, W. C., A. H. Smith and A. Rabe, Retinal Gradient of Outline as a Stimulus for Slant, *Canadian Journal of Psychology 9* (1955), 247-253.

8. Clowes, M. B., On Seeing Things, *Artificial Intelligence 2* (1971), 79-116.

9. Flock, H. R., D. Graves, J. Tenney and B. Stephenson, Slant Judgments of Single Rectangles at a Slant, *Psychonomic Science 7*, 2 (1967), 57-58.

10. Freeman, R. B. J., Function of Cues in the Perceptual Learning of Visual Slant, *Psychological Monographs 80*, 2 (1966).

11. Guzman, A., Computer Recognition of Three-dimensional Objects in a Visual Scene, MAC-Tech. Rep.-59, M.I.T., 1968.

12. Kanade, T., Recovery of Three-dimensional Shape of an Object from a Single View, *Artificial Intelligence 17* (1981), 409-460.

13. Kanade, T. and J. R. Kender, Mapping Image Properties into Shape Constraints: Skewed Symmetry, Affine-Transformable Patterns, and the Shape-from-Texture Paradigm, in *Human and Machine Vision*, Academic Press, 1983.

14. Mackworth, A. K., Interpreting Pictures of Polyhedral Scenes, *Artificial Intelligence 4* (1973), 121-137.

15. Palmer, S. E., The Psychology of Perceptual Organization: A Transformational Approach, in *Human and Machine Vision*, Academic Press, 1983.

16. Stavrianos, B. K., The Relation of Shape Perception to Explicit Judgments of Inclination, *Archives of Psychology*, 1945.

17. Stevens, K. A., Surface Perception from Local Analysis of Texture and Contour, MIT-AI-Tech. Rep.-512, M.I.T., Cambridge, MA, 1980.

18. Weiss, I., 3-D Shape Representation by Contours, *Computer Graphics, Vision, and Image Processing*, to appear.

19. Witkin, A. P., Recovering Surface Shape and Orientation from Texture, *Artificial Intelligence 17* (1981), 17-45.